

OM protein - protein search, using sw model

Run on: July 25, 2003, 17:10:54 ; Search time 30 Seconds

(without alignments)
949.174 Million cell updates/sec

Title: US-09:989-981A-8

Perfect score: 3506
Sequence: 1 MAGKAAEERGLPKGATPGDT.....FMVLVYVSLRFIKORPSQDW 673

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/ptodata1/1/aa/5A_COMB.pcp: *
- 2: /cgn2_6/ptodata1/1/aa/5B_COMB.pcp: *
- 3: /cgn2_6/ptodata1/1/aa/6A_COMB.pcp: *
- 4: /cgn2_6/ptodata1/1/aa/6B_COMB.pcp: *
- 5: /cgn2_6/ptodata1/1/aa/PCTUS_COMB.pcp: *
- 6: /cgn2_6/ptodata1/1/aa/backfilltest.pcp: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	%	Length	DB	ID	Description
1	640.5	18.3	655	4	US-09-345-808-1			Sequence 1, Appl1
2	638.5	18.2	655	4	US-09-767-594-1			Sequence 1, Appl1
3	250	7.1	1684	3	US-08-665-259-25			Sequence 25, Appl1
4	250	7.1	1684	3	US-08-762-500-75			Sequence 25, Appl1
5	250	7.1	1704	3	US-08-762-500-75			Sequence 25, Appl1
6	243	6.9	553	4	US-09-252-991A-21665			Sequence 21665, Appl1
7	242	6.9	330	4	US-09-252-991A-27569			Sequence 27569, Appl1
8	230	6.6	370	4	US-09-352-991A-20719			Sequence 20719, Appl1
9	228.5	6.5	359	4	US-09-328-352-6339			Sequence 6329, Appl1
10	227	6.5	432	4	US-09-328-352-6798			Sequence 6798, Appl1
11	225	6.4	345	4	US-09-352-991A-18872			Sequence 18872, Appl1
12	222.5	6.3	1457	3	US-08-665-259-27			Sequence 27, Appl1
13	222.5	6.3	1457	3	US-08-762-500-27			Sequence 27, Appl1
14	219.5	6.3	336	4	US-08-311-731A-209			Sequence 209, Appl1
15	219.5	6.3	676	4	US-09-252-991A-21751			Sequence 21751, Appl1
16	217.5	6.2	633	4	US-09-198-452A-242			Sequence 242, Appl1
17	217	6.2	285	4	US-09-352-991A-26621			Sequence 26621, Appl1
18	216.5	6.2	345	4	US-09-328-352A-2849			Sequence 3849, Appl1
19	215	6.1	250	4	US-09-328-352-7133			Sequence 7133, Appl1
20	215	6.1	226	4	US-09-314-001C-4600			Sequence 4600, Appl1
21	214.5	6.1	376	2	US-08-997-080-89			Sequence 89, Appl1
22	214.5	6.1	376	2	US-08-997-362-89			Sequence 89, Appl1
23	214.5	6.1	376	3	US-08-873-970-89			Sequence 89, Appl1
24	214.5	6.1	376	3	US-09-095-855-89			Sequence 89, Appl1
25	214.5	6.1	376	4	US-09-324-542-89			Sequence 89, Appl1
26	214.5	6.1	376	4	US-09-205-426-89			Sequence 89, Appl1
27	213.5	6.1	363	4	US-09-352-991A-25307			Sequence 25307, Appl1

28	212	6.0	707	3	US-08-772-270A-4	Sequence 4, Appl
29	212	6.0	707	4	US-09-062-126-4	Sequence 4, Appl
30	211.5	6.0	711	3	US-08-772-270A-12	Sequence 12, Appl
31	210	6.0	929	4	US-09-252-991A-22946	Sequence 22946, A
32	209.5	6.0	217	4	US-09-107-532A-6401	Sequence 6401, A
33	209.5	6.0	402	4	US-09-107-532A-5360	Sequence 5360, A
34	209.5	6.0	402	4	US-09-134-001C-5370	Sequence 5370, A
35	209.5	6.0	460	4	US-09-134-001C-5369	Sequence 5369, A
36	208	5.9	256	4	US-09-107-532A-2308	Sequence 2308, A
37	208	5.9	260	4	US-09-107-532A-5350	Sequence 5350, A
38	207.5	5.9	261	4	US-09-252-991A-33060	Sequence 33060, A
39	207.5	5.9	655	4	US-09-252-991A-18411	Sequence 28411, A
40	207	5.9	227	4	US-09-634-238-289	Sequence 289, A
41	207	5.9	337	4	US-09-134-001C-5550	Sequence 5550, A
42	207	5.9	369	4	US-09-328-352-6505	Sequence 6505, A
43	205.5	5.9	1548	1	US-08-463-092B-7	Sequence 7, Appl
44	205.5	5.9	1548	2	US-08-463-092B-7	Sequence 7, Appl
45	204	5.8	624	2	US-09-252-991A-21625	Sequence 21625, A

ALIGNMENTS

```

RESULT 1
US-09-245-808-1
Sequence 1, Application US/09245808
Patent No. 6313277
GENERAL INFORMATION:
APPLICANT: Doyle, L. Austin
APPLICANT: Abruzzo, Lynne V.
APPLICANT: Ross, Douglas D.
TITLE OF INVENTION: Breast Cancer Resistant
TITLE OF INVENTION: encodes it
FILE REFERENCE: Ross OMD conversion
CURRENT APPLICATION NUMBER: US/09/245,808
CURRENT FILING DATE: 1999-02-05
EARLIER APPLICATION NUMBER: 60/073763
EARLIER FILING DATE: 1998-02-05
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 655
TYPE: PRT
ORGANISM: Human MCF-7/AdvtP cells
US-09-245-808-1

```

Query Match	18.38;	Score 640.5;	DB-4;	Length 655;
Best Local Similarity	27.28;	Pred. No. 3.1e-62;		
Matches 187;	Conservative 139;	Mismatches 273;	Indels 89;	Gaps 21

[illegible]

QY 367 -----DDEDTCESSVTPDITNCLSPTRKPGAVOFTLLIRROISNDRDLPITLLI 419
DB 349 LHQSGGKKKKITVFEKISYTTSTFC-----HQLRWVSKRSFKNLGNPQASIA 397
QY 420 HGAECIMSMITGIFEGHGSIOLEFMDTALLFMGALIPRNVLIDVSKYS----- 473
DB 398 OIIVTVLGLVIGALIFGKNDSTGIONRAGVFL-----TTNOCFFSSVSAVE 446
QY 474 -----BRAMLYELEDGLYTTGPFPAKILGE-LPEHCAYIIYGMPTWLANRPGLOPF 528
DB 447 LFVVEKKLFHIEYISGYRVSSYFLGKLSIDLPMRLPSITICTIVFMGLKPKADAF 506
QY 529 LHFLVLMVVECCRIMALAAALPTFHMAFFSNALYNSFYLAGFMINLSLMTVPA 588
DB 507 FVMEFTLMVAVASASMAIAAGSVSVATLMTICFVFMIFSGLLVNLTTIASWLS 566
QY 589 MISVSEFLMCEGMLKIOFSRRTYKMLGNLT-----IIVSGDKIL--SAMELDSYP 639
DB 567 WLOFSTIPRYGFTALQHNEFLGQNF-CPGLNATGNNPCNATCTGEEYLVKQIDLSPMWG 625
QY 640 LYAIVLYIGLSGGMVLYVSLRPIKQ 667
DB 626 LMKHNVALACMIVITLTAIVLKLFLPK 653

RESULT 2
US-09-767-594-1
Sequence 1, Application US/09767594
Patent No. 6521635

GENERAL INFORMATION:
APPLICANT: Bates, Susan
APPLICANT: Robey, Robert
APPLICANT: The Government of the United States of America
APPLICANT: as represented by the Secretary of the
APPLICANT: Department of Health and Human Services
TITLE OF INVENTION: Inhibition of MXR Transport by Acridine Derivatives
FILE REFERENCE: 015280-402100US
CURRENT FILING DATE: 2001-01-22
PRIOR APPLICATION NUMBER: US 60/177,410
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 655
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: human mitoxanthrone resistance (MXR)/BRCP/ABCP
US-09-767-594-1

Query Match 18.28; Score 638.5; DB 4; Length 655;
Best Local Similarity 27.98; Pred. No. 5.2e-62;
Matches 175; Conservative 131; Mismatches 254; Indels 67; Gaps 17;

QY 80 CONSCELGI-QNLSFKYRSCOMLAIIGSSGGRASLLDVTYGRHGKIKSGQIWINQOP 138
DB 55 CRKPEVEILSNINIGIMKPG-LNAIILPTGGKSSLLDVLAAKRDPSGL-SGDVLINIGAP 112
QY 139 SSPOLVRKC-VAHROHNOQLLNLVRETLAFLAOMELPRTFSQADKREVEDVIAELRL 197
DB 113 RPAHF--KCSNGYVQDDVWGLTVRENLOFSMAHLATVTMTHKENENINRYIEBLGL 170
QY 198 ROCADTRVGMNVYVSGSGRRRSSTVOLAMPGLILDEPSSGDSFAHNILVYTLR 257
DB 171 DKVADSVGTOFTNGVSGGKKRRTSGMELITDPSLIDEPDTGLDSSSTANAVLLKLKR 230
QY 258 LANGNRVLISLHQPSRDIREFDLVILMTSGTPIYGAOHMVQYFTALGYPCPRSNP 317
DB 231 MSKGRITLIFSIIHQPRSRIFKIDPSLTLASGRMLMFGPQOALGYESAGYCEAYINP 290

QY 318 ADFVYDLTSDIR-----SREQLATRE--KAOSLAALFEKVRDL--DQFLKAEK--- 366
DB 291 ADFDILINDGSTAVALNREDFKATLEIESKQDKPLIEKALYVSSYFK-ETKAE 349
QY 367 -----DDEDTCESSVTPDITNCLSPTRKPGAVOFTLLIRROISNDRDLPITLLI 420
DB 350 HOLSGGKKKKITVFEKISYTTSTFC-----HQLRWVSKRSFKNLGNPQASIAQ 398
QY 421 GAEACIMSMITGIFEGHGSIOLEFMDTALLFMGALIPRNVLIDVSKYS----- 473
DB 399 IIVTVLGLVIGALIFGKNDSTGIONRAGVFL-----TTNOCFFSSVSAVE 447
QY 474 -----BRAMLYELEDGLYTTGPFPAKILGE-LPEHCAYIIYGMPTWLANRPGLOPF 529
DB 448 FVVEKKLFHIEYISGYRVSSYFLGKLSIDLPMRLPSITICTIVFMGLKPKADAF 507
QY 530 LHFLVLMVVECCRIMALAAALPTFHMAFFSNALYNSFYLAGFMINLSLMTVPAW 589
DB 508 FVMEFTLMVAVASASMAIAAGSVSVATLMTICFVFMIFSGLLVNLTTIASWLS 567
QY 590 ISKVSFLMCEGMLKIOFSRRTYKMLGNLT-----IIVSGDKIL--SAMELDSYP 640
DB 568 LOYFSTIPRYGFTALQHNEFLGQNF-CPGLNATGNNPCNATCTGEEYLVKQIDLSPMWG 626
QY 641 YAIVLYIGLSGGMVLYVSLRPIKQ 667
DB 627 WKNHVALACMIVITLTAIVLKLFLPK 653

RESULT 3
US-08-665-259-25
Sequence 25, Application US/08665259
Patent No. 6028173

GENERAL INFORMATION:
APPLICANT: Landes, Gregory M.
APPLICANT: Burn, Timothy C.
APPLICANT: Connors, Timothy D.
APPLICANT: Dackowski, William R.
APPLICANT: Van Raay, Terence J.
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:

ADDRESSEE: GENZYME CORPORATION
STREET: One Mountain Road
CITY: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/665,259
FILING DATE: 17-JUN-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Dugan, Deborah A.

REGISTRATION NUMBER: 37,315

REFERENCE/DOCKET NUMBER: IG5-9.1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (508) 872-8400

TELEFAX: (508) 872-5415

INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:

LENGTH: 1684 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-665-259-25

430 TIGELYFGHSIQLSFMDTAALLEMIG---ALIPFNVI-LDVISKCYSERAMLYYELED 484

```

; LENGTH: 330
TYPE: DPM

```

SEQ ID NO 27569
LENGTH: 330

```

RESULT 7
US-09-252-991A-27569
: Sequence 27569, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: ARGGININOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 27569
: LENGTH: 330
: TYPE: PRT

```


Db 27 LDKGLIMQIPSLAQSSLL-----GQ-----NOKVELACHL--LSTLTT 62
QY 71 EKMPWTSPSCNSCEL-----GIQNLSEFKVSGOMLAIIGSSGGRASLDVITG-- 120
Db 63 YKYPEHGPSVLEVVADIKFQHPQALTHIMDKAGEFVSFLPSCGCKTLIRIANGLE 122
QY 121 -REHGKIKSGQWIMNGPSSPOLV---KCAVHRQHNOLLPLTRETFLAFIAQML 175
Db 123 KPDGVYVTKG-----ADITLSEKHC--GIYFQNYALPMLVNEENITFGD--- 170
QY 176 PTFESQARDKREVEDYIAELRLQCADTRYGNMYVNGSGERRYSIGVQLMNPGLI 235
Db 171 KKKMDAQRQSRQVQLLELLELD-----IAKYPNOLSSGQOQVVALARATIPNDLIL 225
QY 236 LDEPTSGDSTFANHLVKTLSRLAKGNRLVLSIHOPRSDIFLFLVILMTSG----- 289
Db 226 LDEPLSALDAQVRLKIRKIRSIQTOLNPTIIVTHDQEALISDRVYVMMNGVLEQID 285
QY 290 TP---YLGAAQHMVYFIAIGPCPRISNPADFYDLSIDRRSREBELTREKASOLA 346
Db 286 TPNHNYKQTOFVAEFIGTMNPLKATCANPNOLEV-LGFIPNLEQCKIKAGENYS--1 342
QY 347 ALELEKVRDLDELKMAETKMDLEDTCVESSVTPIDTNCPLSPTKMGA-----VOQFTT 401
Db 343 GFRPEAVELVDGSGKRESLITL-----PVAVLSTEEFLGAKRRLLCAIHIDIGIOAKH 394
QY 402 LIRQIISND-----FRDLPTLLIH 420
Db 395 LQIEIENTRFKSLQELMFKVNPOLIH 422

RESULT 11

US-09-252-991A-18872
Sequence 18872, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 18872
LENGTH: 345
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18872

Query Match 6.4%; Score 225; DB 4; Length 345;
Best Local Similarity 26.2%; Pred. No. 3.7e-16;
Matches 67; Conservative 53; Mismatches 112; Indels 24; Gaps 5;

QY 88 IONLSFKVSGOMLAIIGSSGGRASLDVITGRHGKIKSGQWIMNGPSSP-----141
Db 31 IQPTLNLITQAGIIFGLIGHSAGKSTLRLI---NLEPSSGRLLVBEEDVATLDAEGL 87
QY 142 QLVKCAVAVRHQNNOLLPLNLTVEFLTAFAOMRLPRTFSQARDKREVEDYIAELRLQCA 201
Db 88 RPRRQVGMIFQHPNLLSSKTVADNIAM--PLRLAGGFSRAVDARVSLARVGLSDHA 145
QY 202 DTRVGNMYVNGSGERRYSIGVQLMNPGLIILDEPTSGDSTFANHLVITLTLAKG 261
Db 146 -----RKYPALSSGQOKQVGIARALACRPSILCDENASALDPTQTASVQLDLAEINNE 200
QY 262 NBLVLISLHOPRSDIFRFLDVLNLTSGPIYLGAAQHMVYFIAIGYCCPRYSNPADRY 321
Db 201 LKLTIVLITLHMDVIRKVCDDVAVNDGAIYDQGVADVFLH-----POHPTTRREV 252

QY 322 VDLTSIDRRSREDELA 337
Db 253 FEAEVRVEDERHDDFA 268

RESULT 12

US-08-665-259-27
Sequence 27, Application US/08665259
Patent No. 6028173
GENERAL INFORMATION:
APPLICANT: Landes, Gregory M.
APPLICANT: Burn, Timothy C.
APPLICANT: Connors, Timothy D.
APPLICANT: Dackowski, William R.
APPLICANT: Van Raay, Terence J.
APPLICANT: Klinget, Katherine W.
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: One Mountain Road
CITY: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665,259
FILING DATE: 17-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dugan, Deborah A.
REGISTRATION NUMBER: 37,315
REFERENCE/DOCKET NUMBER: IG5-9.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 1457 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-665-259-27

Query Match 6.3%; Score 222.5; DB 3; Length 1457;
Best Local Similarity 22.6%; Pred. No. 9e-15;
Matches 78; Conservative 68; Mismatches 138; Indels 61; Gaps 8;

QY 81 ONSCEGLIONLSFKVSGOMLAIIGSSGGRASLDVITGRHGKIKSGQWIMNGPSS 140
Db 22 KMDKRLAKLKLKSLIENOVVSFLGNAGKTTTSLIGL---FPPTSGSATIYCHDIR 78
QY 141 POL--VRKCAVAVRHQNNOLLPLNLTVEFLTAFAOMRLPRTFSQARDKREVEDYIAELRLR 198
Db 79 TEMDEIRKMLGNCPOHNVFLDRLIYVEHIMFTSRL--KSMQDEIRKTDKMIEDLEL- 134
QY 199 QCADRVGNMYVNGSGERRYSIGVQLMNPGLIILDEPTSGDSTFANHLVITLTLAKG 258
Db 135 ---SNKRSLVQTLISGCMKRLISVAIAFVGGSSRAIILDEPTAGVDPYARRAIVDLILKY 190
QY 259 AKGNRLVLSLHOPRSDIFRFLDVLNLTSGPIYLGAAQHMVYFIAIGYCCPRYSNPADRY 310
Db 191 KRGKRTILSTHMDL--LGDRIALISHGKLKCCGSPFLKGAIXDGYRLTLVKOPAE 248
QY 311 -----CPRYSNPADRYVD-----LTSIDRRSR 332

Db 249 PGTSOEPGLASSPGCCPRLSSCSEPVSQFIRKRVASSLTVSDTSTELSYILPSEAVK 308
QY 333 EOEIATREKAOSIALAFLEKVRDLDPLMKRATKCOLDDPCVSS 377
Db 309 AFERLFOGLEHSLDALHLSRGLMDTTLVEVFLKVSSEDSLENS 353

RESULT 13
US-08-762-500-27

: Sequence 27, Application US/08762500
: Patent No. 6030806
: GENERAL INFORMATION:
: APPLICANT: Landes, Gregory M.
: APPLICANT: Burn, Timothy C.
: APPLICANT: Connors, Timothy D.
: APPLICANT: Dackowski, William R.
: APPLICANT: Van Raay, Terence J.
: TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
: TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
: NUMBER OF SEQUENCES: 83
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: GENZYME CORPORATION
: STREET: One Mountain Road
: CITY: Framingham
: STATE: Massachusetts
: COUNTRY: United States of America
: ZIP: 01701
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/762,500
: FILING DATE: 09-DEC-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/665,259
: FILING DATE: 17-JUN-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US96/10469
: FILING DATE: 17-JUN-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Dugan, Deborah A.
: REGISTRATION NUMBER: 37,315
: REFERENCE/DOCKET NUMBER: IGS-9.3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (508) 872-8400
: TELEFAX: (508) 872-5415
: INFORMATION FOR SEQ ID NO: 27:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1457 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: US-08-762-500-27

Query Match 6.3%; Score 222.5; DB 3; Length 1457;
Best Local Similarity 22.6%; Pred. No. 9e-15;
Matches 78; Conservative 68; Mismatches 138; Indels 61; Gaps 8;

QY 81 QNSCEGLONLSKRVSGMLAIISSGCGRASLDVYTRGGRGKSGQIWINQOPSS 140
Db 22 KDKRLAKLKLKSLINTLVENQVVSFLGHNGAGKTTTMSILNGL--FPPTSGSATIYCHDR 78
QY 141 POL--VRGCAVHROHNOILPNTVRETLAFLAQMRLPRTFSQAQDRKVEDYIALRLR 198
Db 79 TENDETRNKILGMCPOHNVFLDRLLVEHLMFYSL--KSMQOEIRKETDKMIEDLEL- 134
QY 199 OCADTRVGNMYVGLSGGERRRVSIGVOLLMPNGLILDEPTSGIDSPFAHNLVKTLRL 258

Db 135 ----SNKRHSLVOTLSSGMRKRLVAIAFVCGSRAIILDEPTAGVDPIARRAINDLI 190
QY 259 AKGNRLVILSHQPSRDIERLVDLLMT-----SGTPPIYGAQHVVQYTAIGYP-- 310
Db 191 KPGRIILSTHMHMEADL--LGDRIALISHGKLKCCGSPFLKGAHXGYRLTLVKQPAE 248
QY 311 -----CPRYSNPADFYVD-----LTSIDRRSR 332
Db 249 PGTSOEPGLASSPGCCPRLSSCSEPVSQFIRKRVASSLTVSDTSTELSYILPSEAVK 308
QY 333 EOEIATREKAOSIALAFLEKVRDLDPLMKRATKCOLDDPCVSS 377
Db 309 AFERLFOGLEHSLDALHLSRGLMDTTLVEVFLKVSSEDSLENS 353

RESULT 14
US-08-311-731A-209

: Sequence 209, Application US/08311731A
: Patent No. 6583266
: GENERAL INFORMATION:
: APPLICANT: SMITH, DOUGLAS
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
: TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
: NUMBER OF SEQUENCES: 411
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: WOLFE, GREENFIELD & SACKS, P.C.
: STREET: 600 ATLANTIC AVENUE
: CITY: BOSTON
: STATE: MASSACHUSETTS
: COUNTRY: USA
: ZIP: 02210
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/311,731A
: FILING DATE:
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: GATES, EDWARD R.
: REGISTRATION NUMBER: 31,616
: REFERENCE/DOCKET NUMBER: C0044/7125
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/720-3500
: TELEFAX: 617/720-2441
: INFORMATION FOR SEQ ID NO: 209:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 396 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHETICAL: YES
: ORIGINAL SOURCE:
: ORGANISM: Mycobacterium leprae
: US-08-311-731A-209

Query Match 6.3%; Score 219.5; DB 4; Length 396;
Best Local Similarity 28.5%; Pred. No. 2e-15;
Matches 69; Conservative 52; Mismatches 84; Indels 37; Gaps 9;

QY 88 IONLSKRVSGMLAIISSGCGRASLDVYTRGGRGKSGQIWINQOPSSPOLVR-K 146
Db 24 VHDLSITVADGEFLILIGSGCGKTTLLNMIAGL---EDISSGELRIDGRVVEKAPKOR 80
QY 147 CVAHVROHNOILPNTVRETLAFLAQMRLPRTFSQAQDRKVEDYIALRLRCAOT 203
Db 81 DIAMVQSYALVHYHMYVRQNIAPFLALAYK-----KAIIQAVSETQIIDLTPDLDR 134
QY 204 RVGNMTVRLGSLGGERRRVSIGVOLLMPNGLILDEPTSGIDSPFAHNLVKTLRLAKGN 263

Db 135 KPSQ-----LSGGGRVAMGRAIVRHAFIMDEPLSNID---AKLRVTTGCEIARLQR 186
QY 264 ---LVLSLHQPDSIDFLEEDVLMTSGTPYILGAQHWOYFALIGYPCPRYSNPADF 320
Db 187 RLGAFTVYVTHDQTEAMTGLDGRVYVVRSG-----VVG---QIGTPDELXYERPVNL 233
QY 321 YV 322
Db 234 FV 235

RESULT 15

US-09-252-991A-21751
Sequence 21751, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 21751
LENGTH: 676
TYPE: PR
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21751

Query Match

6.3%; Score 219.5; DB 4; Length 676;

Best Local Similarity 26.6%; Pred. No. 3e-15; Mismatches 97; Indels 57; Gaps 10;

Matches 75; Conservative 53; Mismatches 97; Indels 57; Gaps 10;

QY 46 TLEVRDL--NYQVDLASQVPMFPEQLAQEFKMPYTPSPSCNCELGIONLSFKYRSGOMLAI 103
Db 293 TELRNLMKSYSGSLADT-----LRDIRLKTIDSGEFLIL 326
QY 104 IGSSCGGRASLDVITGKGHGKIKSGOIWINGQPPSPQVLR-KCVAHYRQHNOLLPLNT 162
Db 327 VGPSCGKSTLMNCIAGL---ESITGAILVDGQDISGSPKRDIAWVFQSYALYPTMS 383
QY 163 VRETLAFLAOMRLPRTFSQARQDRVEDYIAELRLQCADTRVGMAYVAGLSGGERRRYS 222
Db 384 VRENIAFGIKT---RKMPQALIDEVAVARILQIEHLERRPSQ-----LSGGGOORVA 435
QY 223 IGVOLLMPGILILDEPTSGIDSFYAHNLVKTLSRLAKGNRLVLSLHQ--PRSDIFRLF 280
Db 436 MGRALARRPKYVLFDEPLSNID-----AKLRVEMRTELKLMHQRLLKTTVYVTH 484
QY 281 DLVILMTSGTPYILGAQHWOYFALIGYPCPRYSNPADFVY 322
Db 485 DQIEAMTGLDGRVAV-MKDGVIQOF---GTPQOIYNDPANTLFV 522

Search completed: July 25, 2003, 17:16:09
Job time : 33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 17:08:49 ; Search time 43 seconds

(without alignments)
1505.152 Million cell updates/sec

Title: US-09-989-981a-8

Perfect score: 3506

Sequence: 1 MAGKAABERGLPKGATPPDT.....FMVLYVSLRFIKRPSQDW 673

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: PIR_76:**

1: piri:**
2: pir2:**
3: pir3:**
4: pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	735.5	21.0	725	2	C84423
2	723.5	20.6	646	2	C86441
3	700	20.0	609	2	E96742
4	668.5	19.1	737	2	T46101
5	656	18.7	687	1	FYFWM
6	653	18.6	656	2	JC7860
7	653	18.6	1294	2	S77690
8	651.5	18.6	635	2	T08934
9	618	17.6	638	2	G02068
10	595	17.0	687	2	D96553
11	591	16.9	725	2	T47652
12	590	16.8	559	2	B88474
13	589.5	16.8	720	2	T47648
14	586.5	16.7	708	2	T47650
15	579	16.5	658	2	T31958
16	575.5	16.4	678	2	H96552
17	573.5	16.4	646	2	JC7777
18	562.5	16.0	1049	1	S19421
19	561.5	16.0	739	2	T45891
20	561.5	16.0	755	2	G84791
21	556.5	15.9	740	1	T02567
22	550	15.7	649	2	A84509
23	548	15.6	662	2	T47649
24	547	15.6	590	2	B96573
25	539	15.4	695	2	T21109
26	534	15.2	633	2	T19189
27	533.5	15.2	659	2	E86313
28	527	15.0	1443	2	T02491
29	525.5	15.0	577	2	T04229

30	524	14.9	608	2	T34391	hypothetical prote
31	521	14.9	639	2	G88839	protein C1066.5 [1
32	511.5	14.6	547	2	T31543	hypothetical prote
33	511	14.6	610	2	T19333	hypothetical prote
34	506.5	14.4	1423	2	A86289	probable ABC trans
35	480	13.7	1450	2	T45888	ABC transporter-11
36	474	13.5	1451	2	B86286	ABC-type transport
37	467.5	13.3	1420	2	T02644	ATP-binding cass
38	466.5	13.3	1426	2	T30567	probable ABC trans
39	464	13.2	1364	2	S55517	probable ABC trans
40	463.5	13.2	675	1	FYFEB	protein putative A
41	461	13.1	1469	2	H96622	probable ABC trans
42	452	12.9	1435	2	D96693	probable ABC trans
43	446	12.7	1450	2	A84780	probable ABC trans
44	440.5	12.6	1413	2	G84790	probable ABC trans
45	437	12.5	668	2	S55023	brown protein - fr

ALIGNMENTS

QY	75	WTSPSC-----QNSCELGIONLSFKVRSQGMALITSSGCGRASILDVTC-----RG 122	21.0%; Score 735.5; DB 2; Length 725;
Db	72	WRNTTSLDKSKSVRFILKNVSGEAKPGRLLAINGPSGKTYTLINVLAVQLSLSPRL 131	Best Local Similarity 30.0%; Pred. No. 3,6e-50;
QY	123	HGGKIRSGOIMWINGOPSPOLYKRCVAHROHQLPNTLVRETLAFIAQMRPPTFSQA 182	Matches 186; Conservative 123; Mismatches 229; Indels 81; Gaps 15;
Db	132	H-----LSGLELVNGKRPSSKAYK--LAFVROEDLFFSOLTVRRTLSFAELDPELSSAE 185	
QY	183	QNDKREVDYIAELRQCADTRYGNMYVGLSGGERRRVSIGVQLLMNGIILDEPTSG 242	
Db	186	ERDEYVNNLLKRLGLVSCADSCGDAKVGKISGGEKRLSLACELIASVIFADPTTG 245	
QY	243	LDSEFVHNVLKTLRLAKGNRLVLISLHPRSDIFLFDVLVMTGCTPIYIG-AAQHW 301	
Db	246	LDAFOAEKMETLQKLAOGHVTYICIHOPRGSVYAFDYLVLTEGTVVYGPAPKEPL 305	
QY	302	OYFAIGYPCPRYSNPADFYVDLTSIDRSRQDELATRRKAOSIALFLEKVRDLDPLM 361	
Db	306	TYFGNNGFLCPHPVPAEFLADLISDYSSSEYVSSQKRVAVLDAFSOR----- 356	
QY	362	KAEFDLDEDTVESSV---PLDINCLPSPK-----MPGAVOQFTTLIR 405	
Db	357	-----SSSVYATPLS---MKETKNGMRPRRAKAIVERTDMMHOFLLKR 400	
QY	406	QISNDFDLPTLLINGAECALSMWITIGLFGHSGIQLSFMOTALLMFGILIPNVIL 465	
Db	401	AMMAASROGPTNKVRARMSVASAIVTIGSVFMRRKSGQTSIQDRMGL--OVAI---NTAM 456	

RESULT 2
C86441
probable ABC transporter [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision: 02-Mar-2001 #text_change 31-Mar-2001
Accession: C86441

probable ABC transporter [Imported] - Arabidopsis thaliana

Query Match Similarity	20.6%	Score 723.5;	DB 2;	Length 646;
Best Local Similarity	30.6%	Pred. No. 2.7e-49;		
Matches 208; Conservative 119; Mismatches 262;			Indels 91;	Gaps 20;

RESULT 3

proddle ABC transporter F1/M19.11 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)

Query Match	20.0%;	Score 700;	DB 2;	Length 609;
Best Local Similarity	31.8%;	Pred. No. 1.9e-47;		
Matches 210;	Conservative 103;	Mismatches 225;	Indels 122;	Gaps 22

```

OY 76 TSPSCONSCELGT-ONLSFKVRS-----GOMLAIITGSSGCCRASLIDVY 118
Db 1 SINDSCNITKILLGKQKPSDETRSTBERTILSGVTOMISPGEHMAVILGPGSGSKRLTNLV 61
OY 119 TGRHGKGKSGQJWIMINQOPSPSPOLYRKCVAHROHNOJLPHLYTRETAFIAOMLTPRT 178
Db 62 AGRHAGSNLT-7GKILLINDGKITKQTLK-TGFVAODDILLPHLYTRETLVFAALLRPRS 119
OY 179 FSOQOROKRYVDVLAETLRGOCADPRVCNMMVYRGISGGERRVSTGCVOLLNMPGLILDE 238
Db 120 LTRGKYLRAEVSIVSELQTCENTVAVNTRFKGISGGERKVSIAHELLINPSLVLIDE 179
OY 239 PTFSGDSFTAHNLVYKTLRLAKG-NRLVLISLHOPRSDIFRLFDVLMLTSGTPYIYGA 297
Db 180 PTFSGDSFTAHNLVYKTLRLAKGKTYVTSIHOPSSVFQMFDTVILLSEKCKCFVYGGK 239
OY 298 QHMVQYTAIYGPCPRYSNPADFYVDULTSIDRSRECELAATREK---AOSLALALEKVR 354

```



```

Db      240 RDAMAFESVGSFAPFPMNPAFLDLIA--NGVCOTDGTREKRNROTLYTAY----- 292
OY      355 DLDDFLMKAETKDLDEDCVSSVTPDITNCLPSPKMGAVQOFTYLL----- 403
Db      293 ---DRLLAPOVK-----TCIEVSHPPOD--NAFVKTAVNGG--GTTTCATWFSQCLIL 341
OY      404 -----RROISNDFRDLPTLLHGAECALMSMTIFLFGHSGIOLSFMDTALLPMI-- 455
Db      342 HRLKERRHESFD-----LRIPOVVAASITCIGIMWV--HSDYR--DVHDIRGLGLFFPISI 392
OY      456 --GALIPNVILDVISKCYSERAMLYELEDGLTYTGPPYFAKILIGELPEHCAYIIITGM 513
Db      393 FMGVLPSPFNATFTF-----PQERAITRERASGMITLSSTFMAHVAJGSLSMELVLPASFLT 448
OY      514 PTYWLANTRPGLQPELHFLWLVLVVCRCRIMATAAALLPTFHMASFNSALNSFYIA 573
Db      449 FTYMNVYLRPGIVPELTLVLLVLLVLAOGIGLALGAINDAKASTIVYTMALAFVLT 508
OY      574 GGFMINLSSMTVPA--WISKVSFLKRCFEGMLKIQFSRRTYKMPGLNLTAVSGDKIL 630
Db      509 GGYIVN-----KVPSGMVMKRYVSTTFYCYRLVAIOYQ-----SGEELI 548
OY      631 SAMELDSYPLXA-----IYLVIGLSG-----GFWLYYVSLREIK 666
Db      549 RMLGCDSKGKGASATSAAGCRFVEEYIGVGMVTSVGLFLMEFGYRLAYIALRIK 608

```

RESULT 4

```

746101
ABC transporter-like protein - Arabidopsis thaliana
N:Alternate names: protein T25B15.80
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46101
R:Alcaraz, J.P.; Clabault, G.; Cortet, A.; Maché, R.; Mewes, H.W.; Lemcke, K.; Mayer, K.
submitted to the Protein Sequence Database, January 2000
A:Reference number: 223021
A:Accession: T46101
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-737 <ALC>
A:Cross-references: EMBL:AL132972
C:Experimental source: Cultivar Columbia; BAC clone T25B15
C:Genetics:
A:Map position: 3
A:Intons: 122/1; 146/3; 225/2; 277/2; 338/3; 422/2; 535/1; 628/3; 664/3
A:Note: T25B15.80

```

```

Query Match      19.1%; Score 668.5; DB 2; Length 737;
Best Local Similarity 28.0%; Pred. No. 7,7e-45;
Matches 189; Conservative 134; Mismatches 258; Indels 93; Gaps 16;

OY      25 DRLFSSSDNSLYFTYSGOPN---TLEVRDLNVOVDLASQVPMFEQLAQCRKMPSTSCQ 81
Db      120 DLEDEFAATSSVYKFAEPTFPYLFKIDITYKY-----TTRGMT 160
OY      82 NSCELAGION-LSEFVRSGOMALIISSGCRASLLDVITGRGHGKIKSGOIMINQPS 140
Db      161 SSSSEKSLINGISGAVPGEILLALMGPSGSKTILLNLGGRPMONI--GGSVSYNDPYS 219
OY      141 POLVRCAVAVRHQNDLPLVLTAVETLAFTAQNRKLPFTFSQAQRKVEDVIAELRIROC 200
Db      220 KHLKTR-IGFVTODDDVLFPLHETLTYALRLPRTLTEQEKGEQAAASVIGELGERC 278
OY      201 ADRVGMVYRGISGGERRVISGVOLAMPGLILDEPSSGDSFLAHLVYTLSLAK 260
Db      279 QDMIGGSFVGVSGGERRKVCIGNEIMTNPISLLIDEPSSDSTALAKIVMHCIAK 338
OY      261 GNLVLISLHQPASDIFRLFDVLMTSGPIYGAQHVOVFTYIAGYCPRYSNAPD 320
Db      339 AGKTYITTIHOPSSRLFHRDKLVLSRGSLLYFGAASEMASTFSSIGCSPLLAAMNAER 398
OY      321 YVDLTIIDRRSRQELATREKAQSL-AALFLEKVRDLDDFLMKAETKDLDEDCVSSVT 379

```

```

Db      399 LLDLVNGNNDISVPSALKEKKIIRLELYANVK-----CDVEQYILE--AKTQIA 450
OY      380 PLDTNCLPSPKMGAVQOFTLLIRQISNDFRDLPTLLHGAECALMSM----- 429
Db      451 VNEKKILAMPVLDDEVRKIMITCPKREMG-----LSMWEQCLSLRQIKERRHY 501
OY      430 -----TIGLFEGHGSIQLSFMDTALLPMIIGALIPNVILDVISKCYSER 475
Db      502 FEMLVATQVLSAIIIGLIMW--QSDTTSQRPFRSGLLEFVAVGWFEPFETALFFPQER 560
OY      476 AMLYYELEDGLTYTGPPYFAKILIGELPEHCAYIIIGMPTYWLANTRPGLQPELHFLV 535
Db      561 AMLSKERESNMVRLSAIFYAFRTSDLPDLILPELVVYVFMAGLRARAESEFISVLT 620
OY      536 WLIVFCRCRIMATAAALLPTFHMASFNSALNSFYLAGGFMINLSSMTVP--AMISK 592
Db      621 FLCIVAAOGIGLALGASLMDLKAATTLASVVTWTFMAGGYEVK-----KVPETIAMIRF 675
OY      593 VSELMRCFEGMLKIQFSRRTYKMPGLNLTAVSGDKILSAMELDSYPLXAIYVIGLSG 652
Db      676 MSENHYTKLVKVOYE-----ELMESVNGEIESGLK---EVSALVAMIT----- 718
OY      653 GFWLYYVSLREIK 666
Db      719 GYRLVAYFSLRMRK 732

```

RESULT 5

```

FYEFW
white protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 31-Dec-1990 #sequence_revision 17-Feb-1995 #text_change 19-Jan-2001
C:Accession: S08635; S07263; S10240
R:Pepling, M.; Mount, S.M.
Nucleic Acids Res. 18, 1633, 1990
A:Title: Sequence of a cDNA from the Drosophila melanogaster white gene.
A:Reference number: S08635; MUID:90221897; PMID:2109311
A:Accession: S08635
A:Molecule type: mRNA
A:Residues: 1-687 <PEP>
A:Cross-references: EMBL:X51749; NID:98825; PIDN:CAA36038.1; PID:98826
R:O'Hare, K.; Murphy, C.; Lewis, R.; Rubin, G.M.
J. Mol. Biol. 180, 437-455, 1984
A:Title: DNA sequence of the white locus of Drosophila melanogaster.
A:Reference number: S07263; MUID:85134865; PMID:6084717
A:Accession: S07263
A:Molecule type: DNA
A:Residues: 1-24, 'LIFEIPYHCVTAD', 30-334, 'ITLHLSYPAWVPSVLPTRIRKFTYRCNPLCPDGSSPI
A:Cross-references: EMBL:X02974; NID:910873; PIDN:CAA26716.1; PID:910874
A:Experimental source: Strain Canton S
R:O'Hare, K.
submitted to the EMBL Data Library, June 1985
A:Reference number: S10240
A:Accession: S10240
A:Molecule type: DNA
A:Residues: 1-24, 'LIFEIPYHCVTAD', 30-687 <ONAA>
A:Cross-references: EMBL:X02974; NID:910873; PIDN:CAA26716.1; PID:910874
A:Experimental source: Strain Canton S
C:Genetics:
A:Gene: white; w
A:Cross-references: FlyBase:FBgn0003996
A:Intons: 24/3; 116/1; 334/2; 439/3; 483/3
C:Superfamily: fruit fly white protein; ATP-binding cassette homology
C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop; transmembrane protein
F:113-137/Domain: ATP-binding cassette homology <ABC>
F:113-137/Region: nucleotide-binding motif A (P-loop)
F:261-265/Region: nucleotide-binding motif B
F:67,93,472,554,651/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

```

Query Match      18.7%; Score 656; DB 1; Length 687;
Best Local Similarity 30.3%; Pred. No. 6,9e-44;
Matches 178; Conservative 113; Mismatches 265; Indels 32; Gaps 10;

```

```

OY      88 IONLSPFVRSGOMLAIIGSSGGRASLLDVTGKHGC--KIKSGQIWMGQSSPOLVR 145
      113 LKNVCGVAYFGLLAWGSSGACKRTLLNLALFASPGIOVSPSGMRLNGQPDAREMQ 172
OY      146 KCAVAVRQHNOLLPNLTVEETLAFIAOMRPFPSQAOBKRVADVAELRLRCACATRV 205
      173 ARCAVVOODDLFGISLAREHLLFOANVRPRRLTYRQARVQVQIOLSLSCQHTTI 232
OY      206 G-NMYVRLSGGERRRVSIGVOLLMNGILLDEPTSGDSTFANLTKTSLRLAKNRL 264
      233 GVGGRVYKLSGGERRKRLAFASFAITDPEPLICEPTSGDSTFANVVOYLKRLSKQKRT 292
OY      265 VILSLHOPSDIRLRLDVLMTSGPTIYLGAAOHVQYFTALGTPCPKRSNPADFTVDL 324
      293 VILTHOPSELELFDKILLMAEGRAVAFGTPSEADPFSSYGAQCPMYNPADFTVQV 352
OY      325 TSIDRRSRQELATREAGSLALF-LEKY-RDLDPFLMAEKTKDDEDCEVSSVPLD 382
      353 LAY---VPGREISRRIKRICDNFAISKVARDMEQL---ATKMLEK-----PLE 397
OY      383 TNCLESP---TKMPAVOQFTLLIRROISNDFRDLPTLLIHGAECIAMSMTIGFLYFGH 438
      398 -----OPENGYTKATWQFRAVLWMSLVLEKEPLVAVRLIQTMTVAIILGLIFLGQ 452
OY      439 GSTQLSFMDTALLPMIGALIPPNVILDVISKYSEKAMLYELEDEGLYTGPFPAKIL 498
      453 QLTQVGMVINGAIFLELTNMTFQNVFATINFTSELPEVPMREARSRLNCDYFLKTI 512
OY      499 GELPEHCAYIIIIYGMPTYMLANLRPGLOPFLHFLVWLVVPCCRIMALLAALLPFHH 558
      513 AELPELTVPLVYTAIATPMIGALIPPNVILDVISKYSEKAMLYELEDEGLYTGPFPAKIL 572
OY      559 ASEFSNALYNSEFLAGFMINLSLMTVPAMISKVSFLMCEGLMKIOPS---RRTYKM 615
      573 ALSVGPVLIIPFLFGFGLNSGSPVYIKMLSYLSMFRANGLINOMADVEPGEISC 632
OY      616 PLGNLTIANVSGDKILSAMELDSTPLIYALIVIGLSGGENVLYYSILR 663
      633 TSSNTQPSGKRVILETLNFSADPLDLYVGLAILVSEFVLAIALR 680

```

RESULT 6

```

JC7860
brain multidrug resistance protein, BMDP - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 18-Nov-2002 #sequence_revision 18-Nov-2002 #text_change 31-Mar-2003
C:Accession: JC7860
R:Eisenblatter, T.; Gallia, H.J.
Biochem. Biophys. Res. Commun. 293, 1273-1278, 2002
A:Title: A new multidrug resistance protein at the blood-brain barrier.
A:Reference number: JC7860; M0ID:22050127; PMID:12054514
A:Accession: JC7860
A:Molecule type: mRNA
A:Residues: 1-656 <EIS>
A:Cross-references: GB:AJ420927
A:Experimental source: brain
C:Comment: This protein, a new transport protein of the ATP-binding cassette (ABC) super
exclusion of xenobiotics from the brain and participates in drug transport across the bl
C:Genetics:
A:Gene: bmdp

```

```

Query Match      18.6% Score 653; DB 2; Length 656;
Best Local Similarity 27.6%; Pred. No. 1.1e-43;
Matches 192; Conservative 133; Mismatches 270; Indels 100; Gaps 20;

OY      18 OPTSGLODLRFSSSESNLSLYTSGOPNLEVRDLNVOYDVLASQVWFQOLAQFKRPWIS 77
      15 RNTNGL-----PGSSNELKTSAGG--VLSFHDICRYKRVKSGFL-----54
OY      78 PSCQNSCELGI-QNLSFVRSGOMLAIIGSSGGRASLLDVTGTR--GHGCKIKSGQIWI 134
      55 -CRKTVKEKILTNINGIMKPG-LNALIGPTGGKSSLLDVLAAARDPHG-----LSCDVI 108

```

```

OY      135 NGQSSPOLVYKRC-VAHVROHNOPLLNLVRETLAFIAOMRPFPSQAOBKRVEDVIA 193
      109 NGAPRANF--KCMGSGVVODDVVMGLTYRENIQESALRLPTTNNHKNENINAVIQ 166
OY      194 ELRLRCQADRVGCMYVRLSGGERRRVSIGVOLLMNGILLDEPTSGDSTFANLTVK 253
      167 ELGLDKVAADSKVQFOTPRVSGGERRKRTSIAELITDPSILFIDEPTTGDSSFANVLL 226
OY      254 TLSRLAKGNLVIISLHOPSDIRLRLDVLMTSGPTIYLGAAOHVQYFTALGTPCPR 313
      227 LKRMKSQGRITIFSIHQPRYSITPKLPDSITLLASGLHMRGPARALGYFASIGYCEP 286
OY      314 YSNPADFTVDLTS-----IDRRSRQELATREK-----AQSLA-----LFLEK 352
      287 YNPNADPEFLVINDSSAVVLSRDRBGQOPEPEPEKPTPLIDKLAATYTNSSFFKDT 346
OY      353 VRDLDFLMAETKDEDCEVSSVTPDITNCLPSPKMGAVOQFTLLIRROISNDFR 412
      347 KVELDQFSGGRKKR---KSSVYKEVYTTSSFC-----HQLRWISRSGFKMLLG 391
OY      413 DLPTLLIHGAECIAMSMTIGFLYFGHSIQLSFMDTALLPMIGALIPPNVILDVISKY 472
      392 NPOASVQIITVITILGLVIGALIFYDLKNDPSGIONRAGVLEFL-----TTNOCF 440
OY      473 S-----ERAMLYELEDEGLYTGPFPAKILGE-LPEHCAYIIIIYGMPTYMLANL 521
      441 SSVASVAVELVYKELFHEITSGYRRSSIFFGKLSDLLPMRMLSIITTCITYFLGL 500
OY      522 RGLDPLFLHFLVWLVVPCCRIMALLAALLPFHHASFSNALYNSEFLAGFMINLS 581
      501 KPAVGSFFIMFTLMVAVSASSMALAIAGQSVSVATLMTISFPMHIFGLLVNLR 560
OY      582 SLMTVPAMISKVSFLMCEGLMKIOPSRTTYMPLGNLT-----IAVSGDKIL--SA 632
      561 TVVPMLSWLOFYFIPRGFSALQYNELGNF--CPGLNVTNTNCSFALCTGAEYLENQG 619
OY      633 MELDSTPLIYALIVIGLSGFMVLYYSILRFLIKQ 667
      620 ISLSANGIMQNHVALACMAYIFLTINYLKILLIKK 654

```

RESULT 7

```

S77690
Probable membrane protein YOL075c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein O1125; hypothetical protein O1130; hypothet
C:Species: Saccharomyces cerevisiae
C:Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 19-Apr-2002
C:Accession: S77690; S66767; S66768
R:Alexandraki, D.; Katsoulou, C.; Tzermita, M.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S66756
A:Accession: S77690
A:Molecule type: DNA
A:Residues: 1-1294 <ALE>
A:Cross-references: EMBL:Z74816; MIPS:YOL075c
A:Note: this is a revision to the sequence from reference S66756
A:Accession: S66767
A:Molecule type: DNA
A:Residues: 1-179, 'TRTGVFLVYKRED' <ALM>
A:Cross-references: EMBL:Z74816
A:Experimental source: strain S288C
A:Note: this sequence has been revised in reference S77690
A:Note: this was assumed to be protein YOL074c
C:Genetics:
A:Cross-references: SGD:S0005435

```

A:Map position: 15L
 A:Note: Y0L075C
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
 C:Keywords: ATP; nucleotide binding; P-loop; transmembrane protein
 F:45-263/Domain: ATP-binding cassette homology <ABC1>
 F:62-69/Region: nucleotide-binding motif A (P-loop)
 F:376-392/Domain: transmembrane #status predicted <TM1>
 F:469-485/Domain: transmembrane #status predicted <TM2>
 F:496-512/Domain: transmembrane #status predicted <TM3>
 F:606-622/Domain: transmembrane #status predicted <TM4>
 F:700-716/Domain: ATP-binding cassette homology <ABC2>
 F:727-734/Region: nucleotide-binding motif A (P-loop)
 F:1042-1056/Domain: transmembrane #status predicted <TM5>
 F:1125-1141/Domain: transmembrane #status predicted <TM6>
 F:1177-1193/Domain: transmembrane #status predicted <TM7>
 F:1269-1285/Domain: transmembrane #status predicted <TM8>

Query Match 18.6% Score 653; DB 2; Length 1294;
 Best Local Similarity 30.1% Pred. No. 2.9e-43;
 Matches 171; Conservative 111; Mismatches 239; Indels 48; Gaps 13;

QY 88 IONLSPKVRSGOMLAIIGSSGGRASLDVITGRGKGKIKSGQI----- 132
 DB 45 VNTFSMDLPBGSVMAVWVGSGSKTTLNVLASKISGILTHNGSIRVLEDTGSEPNETE 104
 QY 133 ---WINGOPSSPOLVRKCAVHRQHNLPLNLTVEETLAFIAOMRLPRTFSQAORDKR- 187
 DB 105 PKRAHLIDGO-DHPLOKHVIMAYLPQDDVLSPLRTCTETLKFADLKL-----NSSERTKLL 159
 QY 188 -VEDVLAELRLKCCADTRNGNMYRGLSGGERRRISIGVQLMNPGLILDEPTSGLDSE 246
 DB 160 MVBOLLEELGKCCADTLVGDNSHRGLSGGERRRISIGVQLMNPGLILDEPTSGLDSE 219
 QY 247 TAINLVKTLRLAK-GNRVLISLHOPRSDIFRLFDVILMTSGTPIYLGAAGHMYQYF 305
 DB 220 SAEVLVITKLKLKEDGRTFIMSIHQPRSDILFLDQVCLISKGNVYCDKMDNTPIYFE 279
 QY 306 AIGYPCPRYSPNADFYVDLTSIDRSREDELTAREKASLAALFLEKVRDLDFEIKAE 365
 DB 280 SIGYHVPOLVNPADYFDLDSVDSRDKERATQSLNSL-----IDHMD-----YERH 330
 QY 366 KDLDETCVESSSTPLDNLCSPTMGAQVQFTLLIRQISNDRDLPTLLIHGAENK 425
 DB 331 LQDAESYI-SNAETIOIONM--TTRLP-FMKQVTLTRNFKLNSDVYTLSTFAEPL 386
 QY 426 LMSMTIGFLYFGHGSIQLSFMOTALLFMIGALIP--ENVILDVISKCYSERAMLYELE 483
 DB 387 IIGTCGMIYKPRDKSSISGLRTTACLYASTILOCYLVILFDTYRILCEODIALYDERA 446
 QY 484 DGLYTTGPPYFEFA-KILGELPERHCAYIIGMPTMYLANLRPGLOPRLHFLVLMVAVFC 542
 DB 447 EGSVPLAFIVARKISLFLSDPAMTIVISITIFHFGLEADARKFFYPFAVEFLQOLSC 506
 QY 543 RIMATAAALALPETHMASFFSNALYNSFYLAGFMINISLMTVPAMISKVSFLRMCFCG 602
 DB 507 SGLSMISVAVSRDPSKASLVGNMTFTVLSMCGGFNAKVMPIYVMIKIAFTWVSFQT 566
 QY 603 LKLIQPSRR-----TYKMLGNITIVSG 626
 DB 567 LMSSTFTNSYCTTNDLDECLAGNOILEVYG 595

RESULT 8
 T08934
 hypothetical protein F27G19.20 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 17-Mar-2000
 C:Accession: T08934
 R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, A.
 submitted to the Protein Sequence Database, May 1999
 A:Reference number: Z16519
 A:Accession: T08934
 A:Molecule type: DNA

A:Residues: 1-635 <BNV>
 A:Cross-references: EMBL:AL078467; GSPDB:GN00062; ATSP:F27G19.20
 A:Experimental source: cultivar Columbia; BAC clone F27G19
 C:Genetics:
 A:Gene: ATSP:F27G19.20
 A:Map position: 4
 A:Introns: 38/3; 253/L; 304/L; 414/3
 C:Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 18.6% Score 651.5; DB 2; Length 635;
 Best Local Similarity 31.1% Pred. No. 1.4e-43;
 Matches 191; Conservative 104; Mismatches 240; Indels 79; Gaps 19;

QY 30 SESNSLXYFY----SGQPNLEVRDLNLYQVLAQVPEFOLAQFKMPTSPSCNSCEL 86
 DB 17 TNDRSILPFSIFKANNPVLKRENLVYTVKLDQSGCF-----GKNDKTEERT 65
 QY 87 GIONLSPKVRSGOMLAIIGSSGGRASLDVITGRGKGKIK-SGQIWINGOPSSPOLVR 145
 DB 66 ILKGLTIVPGEILAMLPSSGSKTSLTALGVRGEGKGLTGNISYNNKPLS-KAVK 124
 QY 146 KCAVHRQHNLPLNLTVEETLAFIAOMRLPRTFSQAORDKRVEDVIAELRLKCCADTRV 205
 DB 125 RTTGFEVYQDDALVPLNLTVEETLAFIALRLPNSFKQEKIKQAKAVTELDGRCDTII 184
 QY 206 GNNVYRGLSGGERRRISIGVQLMNPGLILDEPTSGLDSEFTAHNLVYKTLRLAKGNRV 265
 DB 185 GGFELKGVSGGERRRISIGVQLMNPGLILDEPTSGLDSEFTAHNLVYKTLRLAKGNRV 244
 QY 266 LISLHOPRSDIFRLFDVILMTSGTPIYLGAAGHMYQYFAIGY-PCPRSNADPVDL 324
 DB 245 VTIIRHDP-----SKGNPYFELGSNADYFASVGSPLVERINSSDFLDI 290
 QY 325 TSIDR-----RSRDELTAREKASLAALFLEKVRDLDFEIKAEKTDLDEDTG---V 374
 DB 291 ANCKPLLVISCMWVSQDSQREPKAKALVATYKTNLDSVINEKGD---DLCKNR 347
 QY 375 ESSVTPLDNLCS-PTKMGAVQVFTLLIRQIS-----NDRDLPTLLIHGAENKISM 429
 DB 348 ESS-RVANTYTDMDPTW---WQDFCVLLKRLKORRHDSFGMKV-----AQIFVSP 397
 QY 430 TIGFLYFGHGSIQLS-FMDPALLFMIGALIPENVILDVISKCYSERAMLYELEGLT 488
 DB 398 LGLLW---QYKISRLQDQIGLFTISSWAFPLPQOFTTPPORAMLOKRRSSGMR 454
 QY 489 TGPYFPAKILGELPERHCAYIIGMPTMYLANLRPGLOPRLHFLVLMVAVFCRIMALA 548
 DB 455 LSPYFLSRVYGDLPMELILPCTCLVITTYMAGLNHLANFEVLLVLAHVAVSGGLGLA 514
 QY 549 AALALPETHMASFFSNALYNSFYLAGFMINISLMTVPAMISKVSFLRMCFCGILKIOF 608
 DB 515 LGLVMDOKSATFLGSVIMLFTLLAGGYVQ-----HVPFISIMKY-----VSI 559
 QY 609 SRRTY-MLPGNT 621
 DB 560 GYTYKLILDGYYT 573

RESULT 9
 G02068
 white homolog - human
 C:Species: Homo sapiens (man)
 C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 02-Feb-2001
 C:Accession: G02068
 R:Croop, J.M.; Tiller, G.; Fletcher, J.A.; Lux, M.; Raab, E.; Goldenson, D.; Arcin
 submitted to the EMBL Data Library, August 1995
 A:Reference number: H00769
 A:Accession: G02068
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-638 <CRO>
 A:Cross-references: EMBL:U34919; NID:g1314276; PIDN:AACS1098.1; PID:g1314277
 C:Genetics:

A:Gene: white
 C:Superfamily: fruit fly white protein; ATP-binding cassette homology
 C:Keywords: ATP; nucleotide binding; P-loop
 F:51-253/Domain: ATP-binding cassette homology <ABC>
 F:78-85/Region: nucleotide-binding motif A (P-loop)

Query Match 17.6%; Score 618; DB 2; Length 638;
 Best Local Similarity 25.7%; Pred. No. 6.5e-41;

Matches 173; Conservative 130; Mismatches 266; Indels 104; Gaps 18;

```

OY 33 DNSLYFT--YSGOPN-----TLEVRDLNVOVDLASQVPEQOLAQFMPTSPSCQNSCEL 86
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 17 DNNLTFAORFSSLPRAANVIEFRDLSYV---PECPMKRKGYKTL----- 60
OY 87 GIQNSLSEKVRSGOMLAIIGSSCGRASLDVITGRGHGKIRKSGQIWMINGOPSSPOLVRK 146
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 61 -LKGISGFENSELVAIMPGSGAKSTLNNILAGYRETG--MKGAVLINGLPRDLNCFRK 117
OY 147 CVAHROHNOILNLTVRETLAFIAOMRLPRTSQAROKRVEDVIAELRLROCADPVRG 206
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 118 VSCYIMQDMLLPHLTVOEAMVVAHKLQEQ--KDEGRREMYKEILTALGLISCANTFRG 175
OY 207 NMYVGLSGGERRRYSIGVOLLNMPGILLDEPTSGDSTFTHNLVTKTSLRLAKGRVL 266
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 176 S-----LSGGRKRLAIALIELVNNPPVWFDEPTSGDLSASCPQVVSIMKGLAQGRSTI 230
OY 267 ISLHOPRSDIFRLFDVLMTSGTPYILGAQHMYQFYAIGYPCPRYSNPADFYDLTS 326
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 231 CIHQPSAKLFELFDQVLYLSGGOCYRKGVCNIVLYRLDGLNCTYINNPADFYEVAS 290
OY 327 IDRRSREQL--ATREKAOSLALFLKRYDL-----DDELMKAET-----KDL 369
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 291 GEGYDONSRLVRAVRE--GMCDSDHRRDLGDAEVNPFMHPRSEEVKQTKRLKGLR 345
OY 370 EITCVESSTYPLDTNCLPSPFKMGAVOQFTLLIRQISNDRDLPTLLIHGAELCMGM 429
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 346 KDSSEMEGCHSASCL-----TQFCILFKETYSIMRDSVTLHRTSHIGTL 395
OY 430 TIGFLPYFGHSIQLSFMDTALLF-----MIGALIPFNVIIDVISKYSEBAMLYELE 483
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 396 LIGLLYLGIGNEAKVYLSNSGFLPFSMLFLMFAALMP-----TVLTFPLE 440
OY 484 DGL-----YTGPFYPAKILGELPEHCAYIIYIGMPTYWLANLRGLOPFLHFL 534
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 441 MGVFLREHLNWTYSKAYVLAATMADVPQIMFPVAYCSIVYMTSOPSDAVALFYFLAL 500
OY 535 VLVVFFCCRIALAAALLPFHMASEFNSNALYNSFYLAGFMINLSLMTWPMANYSKVS 594
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 501 GFMISLVAVOSLGLLGAASSTSLQVATFVGPVTAIVLVLFSGFFVSDTIPYLOMMASTIS 560
OY 595 FLRWCFEGSLMKIQF--SRRTYKMPGLNLTIAVSGDKILSAMELDSYPLTAIVLYIGLSG 652
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 561 YVRGFEGVILSIYGLREDLHCDIDENCHFKSEAILRELDEVENAKLY-LDFIVLG--- 616
OY 653 GFMVLYVYSIRFI 665
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 617 ----LFFISLRLL 625

```

RESULT 10 D96553

hypothetical protein F5D21.6 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001

C:Accession: D96553

R:Theologos, A.; Becker, J.R.; Palm, C.J.; Federjeli, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cleasby, T.H.; Dewar, K.;

Nansen, N.F.; Hughes, B.; Hultar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;

C.A.; Li, J.H.; Li, Y.; Lin, S.X.; Liu, Z.A.; Lueros, J.S.; Mail, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; PMID:21016719; PMID:11130712

A:Accession: D96553

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-687 <STO>

A:Cross-references: GB:AEO05173; NID:g10092349; PIDN:AAG12758.1; GSPDB:GN00141

A:Gene(s):

C:Superfamily: 1

C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1;

Query Match 17.0%; Score 595; DB 2; Length 687;
 Best Local Similarity 26.9%; Pred. No. 4.9e-39;
 Matches 174; Conservative 127; Mismatches 247; Indels 98; Gaps 19;

```

OY 88 IONLSFKVRSGOMLAIIGSSCGRASLDVITGRGHGKIRKSGQIWMINGOPSSPOLVRK 147
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 45 LDGLNHAEPRGIRIMAIMPGSGSGKSTLDSLAGRLARNIYMTGNLLNKA--RLDYCL 102
OY 148 VAHVROHNOILNLTVRETLAFIAOMRLPRTSQAROKRVEDVIAELRLROCADPVRG 207
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 103 VAVTYQEDILMGTLVVRETIYSALRLSSDLTKREVNIVGTTIELGLDCCADVIGN 162
OY 208 MYVGLSGGERRRYSIGVOLLNMPGILLDEPTSGDSTFTHNLVTKTSLRLAK--GNRLV 266
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 163 MHSRVSSEGERRRYSVALEILRPOILFDEPTSGDLSASAFVIALNLRADGRITV 222
OY 267 ISLHOPRSDIFRLFDVLMTSGTPYILGAQHMYQFYAIGYPCPRYSNPADFYDLTS 326
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 223 SSIHOPSSVFALFDLFLSSGETYVFGESKFAVEFEAGPCPKRNPDSHFLRCIN 282
OY 327 ID-----RSRDELATREKAOSLALFLKRYDLDF-----LMKAETDOLDED 371
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 283 SDFDTVTATLKGSRIRREP--ATSDPLMLATSEI--KARLVENRYSYAKSAKSRIREL 340
OY 372 TCVE-----SSVPLDTNCLPSPFKMGAVOQFTLLIRQISNDRDLPTLLIHGAELCMGM 429
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 341 ASIEGHGHEVKKSGEAT-----WFKQLRTLTKRSFVNMCKRIGYWSRI 385
OY 418 LIHGAECLMSMTIGFLYF--GHGSIQLSFMDTALLFPMGALIPFNVIIDVISKYSEBAMLYELE 483
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 386 VY-----IVSFCVGTIFYDVGH-----SYSLIARVSCGGRTITGFMTMSIGGPPSFE 436
OY 474 ERMALYELDOSTYTGTFYFAKILGELPEHCAYIIYIGMPTYWLANLRGLOPFLHFL 533
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 437 EAKVEYKERLSGYGVSVYIISNVSSPFLVALALITGSIYNNVKKRPVSHAFCL 496
OY 534 LVMLVFFCCRIALAAALLPFHMASEFNSNALYNSFYLAGFMINLSL-----WTVPAM 589
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 497 NIFFSVSYIESLMMVAVLVPNFMGLTIGAGIIGIIMTSGFFLLDLPVFRYP-- 554
OY 590 ISKSFLLWCFEGSLMKIQFSRRTYKMPGLNLTIAVSGDKILSAMELDSYPLTAIVLYIGLSG 652
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 555 ISFMSYGMALQAGAKNDFGLIED--PMFAGEPKMTGQVINKIGQVYTHSKMIDLSAI 613
OY 644 YLIVTGLSGFMVLYVSLRF-----IKQKS 670
   |||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 614 VLIV-----CYRIILFIVLKLKERAPALKAIOAKRYKSLKRPS 655

```

RESULT 11 T47652

ABC transporter-like protein - Arabidopsis thaliana

N:Alternate names: protein T26112.10

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000

C:Accession: T47652

R:Monfort, A.; Casacuberta, E.; Puigdomenech, P.; Mewes, H.W.; Lemcke, K.; Mayer, K.

submitted to the Protein Sequence Database, February 2000

A:Reference number: 224471

A:Accession: T47652

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-725 <MON>
 A:Cross-references: EMBL:AL132954
 A:Experimental source: cultivar Columbia; BAC clone T26112
 C:Genetics:
 A:Map position: 3
 A:Note: T26112.10
 C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12U6.1; ATP-B

Query Match 16.9% Score 591; DB 2; Length 725;
 Best Local Similarity 25.5%; Pred. No. 1.1e-38;
 Matches 175; Conservative 135; Mismatches 273; Indels 104; Gaps 18;

```

QY 44 PRTLEVRDLYNQVQLASQVPMFEQLAQFKMPWTSPSCONCELDIONLSFVRSGOMLAI 103
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 70 PVLNFMNLYQDYTLRRRPF-----SKONGVKILLDVGSGASDGLAV 115

QY 104 IGSAGGRASLLDVTYGRGHGKIKSGQIMINQGP-SSPOLVRKCAVHROHQLPMLT 162
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116 IGSAGGRASLLDVTYGRGHGKIKSGQIMINQGP-SSPOLVRKCAVHROHQLPMLT 174

QY 163 VRELIATIAOMRLPRTSOAQRKRVDAELRLROCADTRGNMNTVRLSGGERRVS 222
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 175 VKETLMFASERFLPRSLSKRKRERBALIDOLGRNAANTVIGDEGHRGSGERRVS 234

QY 223 IGVOILMNPGLILDEPTSGLDSTFAHNLVKTSLRKLAKGNRLVLSLHOPRSDIFRLD 262
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 235 IGDIIDHPIVLFDEPTSGLDSTFAHNLVKTSLRKLAKGNRLVLSLHOPRSDIFRLD 294

QY 283 VILMTSGTPYILGAQHMVOYFALIGYPCPRYNPDAFYDLTSIDRSREOELATREKA 342
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 295 LILSRKSVFNCSPLSPGFDFGRPIPERENISFALDIV-----RELE-GSNEG 347

QY 343 OSIALLFLEKVRDLDFIMK-----AEKROLED-----TCVSSVTP---LDT 383
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 348 KALVD-FNEK-----WOONKISLQSAPOFNKLDODRSLSLKEAINASVSRGKLYSG 398

QY 384 NCLPSPTKM-----PGAVOFTTLIRROISNDPDLPTLLIHGAELAMTIGELY 435
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 399 SSNSNPTSMETVSSYANPLSETFE-ILAKRYMKMNR-MEVL--GTRIVTVVTCCLLA 454

QY 436 FCGISQLSFMPTAALLFMIGALIP--FNVLIDVISKYSERAMLYELEDGLYTTGPYE 493
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 455 TVYMKLDHPRGAGERLTLEAFVYPTMFYCCLDNVFVIOERITFLRETHNMYKRISIV 514

QY 494 FATHIIGELPRHCAYIIITGMPTWLANLRGLOPFLHLFLVWLVCCRIMALAAAIL 553
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 515 ISHLSVSLPOLLAPSLVFSALTFMTVGLSGLEGFEFYCLLIYASFWSGSSVYTFISGV 574

QY 554 PFEHMASFFSNALYNSFYLAGFMINLSL---WTPAMISKYFRLRCEGLMKQFES- 609
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 575 PNMCLCYMSTIYLANCLLSGTYVNRDRIPFTWT--WFHYISILKYPEAVALINEFDD 631

QY 610 -----RRTYKMPGLNLTIAVSGDKIISAMELDS 637
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 632 PSRCFVRGYOVFDSTLLGGVSDSGKVKLETLSTLRKITESTCLRTGSDLLAQGCIQ 691

QY 638 YPLIAYILVIGISGGMVLYVLSRF 664
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 692 LSKMCLMTWTFASGLFEFRILFYFALLF 718

```

RESULT 12

B88474
 protein C05D10.3 [imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 15-Jun-2001
 C:Accession: B88474
 R:Anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998
 A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
 A:Reference number: A75000; MUID:99069613; PMID:9851916

A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Project
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 19
 A:Accession: B88474
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-559 <STO>
 A:Cross-references: GB:chr.III; PIDN:AAA20989.1; PID:9532111; GSPDB:GN00021; CESP:
 A:Note: similar to D. melanogaster white protein
 C:Genetics:
 A:Gene: C05D10.3
 A:Map position: 3
 C:Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 16.8% Score 590; DB 2; Length 559;
 Best Local Similarity 29.3%; Pred. No. 9.1e-39;
 Matches 159; Conservative 98; Mismatches 231; Indels 54; Gaps 11;

```

QY 88 IONLSFVRSGOMLAIIGSSGGRASLLDVTYGRGHGKIKSGQIMINQGPSSPOLVRK 147
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10 LHMVSGAESGKLALAILSSGAGKTYTMNLNLDVQSGILIDGRANKMKITREM 69

QY 148 VAHVROHQLPMLTVRETLAFIAOMRL-PRFESOQRKRVEDVIAELRLROCADPRVG 206
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 70 SAFVOQHDMFVGTMTAREHLOFARLRLMGDOYSDHEROLRVEQVLTOMGLKCADVIG 129

QY 207 -NMVKGLSGGERRRRSIGVOLMNPGLILDEPTSGLDSTFAHNLVKTSLRKLAKGNRLV 265
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 130 IPNOLKGLSCGKKRKLFSASEIILCPKILFCDEPTSGLDAPFMAGHYVALRSLADNMTV 189

QY 266 LISLHOPRSDIFRLDLYLMTSGTPYILGAQHMVOYFALIGYPCPRYNPDAFYDLT 325
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 190 IITIHOPSSHVYLSFNVCCLMACGRVYTLPGDOAPLFEKCGPCPAYVNPADHLIRTL 249

QY 326 SIDRSREDELATREKASIALAFLKVRDLDFIMKAEKRDLEDDEC-----VES 376
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 250 AVIDSRATSMKT-----ISKIR--QGFL-----STDGQSVLAIAGNANKLRAS 292

QY 377 SYTPDLTNCPLSPTKM-----PGAVOFTTLIRROISNDPDLPTLLIHGAELAMTIGELY 435
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 293 FVTGSPTS---EKTKFEFNODYNASFEOTPLAFMWSMLVIRDPNLSVRLQIITAF 349

QY 430 TIGELFEGHSQISLSEMDPALLFMIGALIPENVIIDY-----ISKYSRAMLTYE 481
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 350 ITGIVFF-----OTPVTPATITISNGIM-FNHIRNMFMLOFPNVPVITABLPVLR 401

QY 482 LEDGLYTYGPFPAKILGELPRHCAYIIITGMPTWLANLRGLOPFLHLFLVWLVVRC 541
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 402 NANGVYRTASVFLAKNIAELPOYIILPIILNTIYVNSGLPFWNYCRASLVITILITNV 461

QY 542 CRIMALAALALTFEPMASFFSNALYNSFYLAGFMINLSLWTPAMISKYFRLRCE 601
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 462 AISISYAIVATIEANTDVAMTILPIFYVPIMAFGFFITDPAISYKMLSSLSYFRYGE 521

QY 602 GL 603
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 522 AL 523

```

RESULT 13

T47648

ABC transporter-like protein - Arabidopsis thaliana
 N:Alternate names: protein T15C9.80

C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000

R:Meves, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
 submitted to the Protein Sequence Database, April 2000

A:Reference number: Z24470
 A:Accession: T47648
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-720 <MEM>
 A:Cross-references: EMBL:AL132970

A:Cross-references: EMBL:AL132970
A:Experimental source: cultivar Columbia; BMC clone T15C9
C:Genetics:
A:Map position: 3
A:Note: T15C9.110
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12u6.1;

Query Match	16.7%;	Score 586.5;	DB 2;	Length 708;
Best Local Similarity	26.8%;	Pred. No. 2.4e-38;		
Matches 187;	Conservative 124;	Mismatches 254;	Indels 133;	Gaps 20

QY 44 PNTLEVRDNTAVQLASOVPAWEEOIAQKRMPTSPSCNSECIGLONISFVRSGOMLAV 103
Db 60 PELLTFNNLSTNVYLRRF-----DESRKTA-----SYKTLDDTTGCRDEETLAV 107
QY 104 IGSSGCRASLLDVTYGRGHGKIKSGOIMWNGOP-SSPOLVRKCAVHRHNOJLPNIT 162
Db 108 LGSGGACKSTLIDLACIAGVADSCT-GTVTLNKEKVLQSLRLKVIASAVWQDILLFPHLT 166
QY 163 VRELLAFIAQRLRPTSSQAKORKEVEVIAELRQCADPTRVGMVYRGISGGERRRYS 222
Db 167 VKETLMTFRSERLRPLSPKSKMEKEVELIDLQIGRNADVPYIGEGHVRSGGGERRRYS 226
QY 223 IGVOLLNPNGLILIDEPSTGIDSEFPAHLVYTLSTLAKGNRLVLSLQHPSDIRFLDL 282
Db 227 IGIDITHDPLILFDEPTSGIDSTNAFNVQVLRKIQSSGVVIMSHPQSARITGLDR 286
QY 283 VLMTSGTPYILGAQHVQYF7AIGPCPRYSNPADFYVDLTSDIRSRBOELATREKA 342
Db 287 LIIISHKSVFNKSPVSLPSEFFSGRPPIKENTEPTEFALVY-----RELEGSS----- 336
QY 343 OSLAALFLEKVRDddf--LW-----KAERK-----DLEDPCV-----BESVPL 381
Db 337 -----EGRDIVERNEKQOONOTARATTQSRVSLKEAIIAASVSGKLYSGSSGANPT 388
QY 382 DTNCLPSPTKPGAVQOFTLLIRROISNDFPD-----LPTLLHGAECIMSWTI--- 431
Db 389 SMETVSYVANDP--LAETFIILAKRYIKIMWITPELIGRICITVMTG---LLATVYMR 442
QY 432 -----GFLYFGHGSIGLSPMDTALLEMIGALIFPNVILDYISKYCSBRALY 479
Db 443 LDNTPRGAQRMGFARFGMSTM-----FYCCADNIPVFIDORYFL 483
QY 480 YELEDGLYTGPFYFAKILGELPEHCAYIIITYGPTWLANLRGOLPFLHFTLLVWLAV 539
Db 484 RETTHNAYFRSSSYISHALVSLPQLMLSLIAFAATTEVTGVLSGLBSEFYCYLIIYAF 543
QY 540 FCCRIMALAAALPTFHMAFSEFNALYNSYLAGGFMINLSSLMYPAWISKYSFLMC 599
Db 544 WSGSSTYTFISGLIPNVMMSIMWYIAALSYCLLGLGYINDRIPLYWIMPHYSILKYP 603
QY 600 FEGIMKIQF---SR---RTYKMPIGNLTIANS-----GDKTILSAMELDSTY 638
Db 604 YEAVLINEFPDSCFCFKGVQVFPCTGILLAEVSHMKYKLLDTLSGSLGTRKITESTCLATG 663
QY 639 P-----LVAIYIYIGLSCG--FMWLYYSLAF 664
Db 664 PDLLMOGIGTOLSKMDCIMTLTLAGLEFRILIFYSLSLF 701

```

RESULT 15
T31958
hypothetical protein F02E11.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 31-Jan-2000
C:Accession: T31958
R:Favell, A.; Scheet, P. submitted to the EMBL Data Library, July 1997
A:Description: The sequence of C. elegans cosmid F02E11.
A:Reference number: 221104
A:Accession: T31958
A:Status: preliminary; translated from GB/EMBL/DBJ
M:Molecule type: DNA

```

A:Residues: 1-658 <FAV>

A:Cross-references: EMBL:AF016661; PIDN:AAB6050.1; GSPDB:GN00020; CESP:F02E11.1

A:Experimental source: strain Bristol N2; clone F02E11

C:Genetics:

A:Gene: CESP:F02E11.1

A:Map position: 2

A:Introns: 115/3; 158/3; 214/3; 330/3; 368/2; 448/3; 525/1

C:Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match

16.5%; Score 579; DB 2; Length 658;

Best Local Similarity 28.3%; Pred. No. 8.5e-38;

Matches 180; Conservative 113; Mismatches 257; Indels 86; Gaps 22;

```
QY 78 PSCNSELGIQNTLSPKV---RSGOMLAIIGSSGGRASLDVYITGRHGKIKSGQIW 133
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 62 PECLAVCALPTSSYQISVSGVAPGEVTLALMGSGAGKTLMLNLAHDTNGVEYLGDT 121
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 134 INGPSSPOLYRKCAVAVRQHNOLLPNITVETLAFLAQMRLPRTFSQARDKNEVDYA 193
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 122 VNGKKITKOKROMCAVYQVDLFCGLTYREOLTYAHMRKKNATVQ-QKEREVENVLR 180
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 194 ELRLQCADTRVG-NATVRGLSGERRRVSIGVOLMNPGLILDEPTSGDSTFANLV 252
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 DMNLTDCQNTLIGIPNPKKSTIGKKRIACELTDPKILFCDEPTSGIDAFMASEV 240
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 253 KLSRLAKGNELVLSHOPRSDFRLDLVLTMTSGTPYIYGAQHWOYFTAIQ--YP 310
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 RALIDLANKGKTIIVLHOPSSYFERFMHVCFATKTYHGAVDRLCPFDKLGDFR 300
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 311 CPRSNPADYVDTLSDRSREDELATR-----EKAQSLALFLEKVRD-LDDEL 362
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 301 VPESYNPADVMSSEISISPET-BOEDVTRIEYLIEHYQNSDIGNMLKRTAVDERG 359
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 363 AETKDEDQTCVSSVTPDPTNCLPPTKMPGANVQPTTLIRQISNDFRDLPTLLHGA 422
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 360 GDEDDGESRYNSTFGT-----OFEILKRSLRTTFRDPDLLRVFA 401
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 423 EACIMNTIGFLYFGHSIQI---SFMDTALLFMIGALIPFNVLIDVISKCYSERAM 479
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 402 QILATAILVGIY---NMRYELKQPTIONLEGVAMNCARDMTFLFPFSVANVTISE 458
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 480 YELEDGLYTGYPYFAKILGELPEHCAYIIITGMPTWLANLRGLOPFLH---FL 535
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 459 REHKSNIYSVEAYFLAKSLAELPQYTLPMYGTIIYMAGLVAVTSFLVFPVCIT 518
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 536 MLYVFCGRIMA--LAAALLPTFMASFEFNALYNSFYLAGFMINLSLMTVPAMISK 593
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 519 WYAVSTAYVAGACIEGDEGLVTF-MPEFVLPML-----VFGFVYVANS--IP 569
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 594 SFLRMC---FEGIMKIQFSRRTYKM-----PLGNLT---AVSGDKILSAMELD 639
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 570 SFVSMFKHGFEALEANQW--KEIDKISGCDLINPLNMTTGYCPASDQGITLRG 627
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 640 LYAIYLI-----VIGLSGFVNLVYVSLRPIK 666
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 628 LYANVLILFMSFFVYRIIGL-----VALKIRVRFK 658
```

Search completed: July 25, 2003, 17:15:32

Job time : 46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 17:08:24 ; Search time 110 Seconds

(Without alignments)
1578.813 Million cell updates/sec

Title: US-09-989-981a-8

Perfect score: 3506
Sequence: 1 MAGRAERGLPKGATPQDT.....FMVLVYVSLRFIKQKPSQDW 673

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_23:*

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organelle:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp.rodent:*
- 12: sp.virus:*
- 13: sp.vertibrate:*
- 14: sp.unclassified:*
- 15: sp.rvivirus:*
- 16: sp.bacteriophage:*
- 17: sp.archae:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2871	81.9	673	11	08R543
2	2835.5	80.9	672	11	08C105
3	742	21.2	668	10	09AR04
4	739.5	21.1	672	10	09L182
5	735.5	21.0	725	10	09Z035
6	735.5	21.0	725	10	09ASR9
7	730.5	20.8	648	10	09G6W5
8	723.5	20.6	646	10	09G6R7
9	713	20.3	652	11	08C104
10	709	20.2	662	10	0949Y4
11	700	20.0	609	10	09G8W6
12	695.5	19.8	801	5	08T691
13	668.5	19.1	737	10	09PT51
14	666	19.0	657	11	09R004
15	665	19.0	687	5	09NH94
16	659	18.8	751	10	093YS4

17	658	18.8	687	5	094960	094960 drosophila
18	653	18.6	656	6	08M193	08M193 sus scrofa
19	651.5	18.6	635	10	09SZR9	09SZR9 arabidopsis
20	647.5	18.5	679	5	08IS30	08IS30 bactrocera
21	643	18.3	670	5	077423	077423 bactrocera
22	642.5	18.3	655	4	0967A8	0967A8 homo sapien
23	642.5	18.3	655	4	08IX16	08IX16 homo sapien
24	642.5	18.3	679	5	09BH97	09BH97 ceratilis c
25	634.5	18.1	655	4	09BLD6	09BLD6 homo sapien
26	621.5	17.7	666	11	09F617	09F617 arabidopsis
27	621	17.6	666	11	09EPG9	09EPG9 rattus norv
28	617.5	17.6	669	5	08WR2	08WR2 rattus norv
29	616.5	17.6	691	10	08RW19	08RW19 arabidopsis
30	613	17.5	594	10	09LGC3	09LGC3 arabidopsis
31	612.5	17.5	669	5	08WR1	08WR1 arabidopsis
32	610.5	17.4	609	5	09VQ4	09VQ4 arabidopsis
33	607.5	17.3	703	10	08RXN0	08RXN0 arabidopsis
34	598.5	17.1	785	4	096L76	096L76 homo sapien
35	597.5	17.0	692	5	P91892	P91892 aedes aegypt
36	595	17.0	687	10	09C8K2	09C8K2 aedes aegypt
37	595	17.0	723	10	08LMT5	08LMT5 oryza sativ
38	591	16.9	725	10	09M3D6	09M3D6 arabidopsis
39	590.5	16.8	610	5	P90746	P90746 caenorhabdi
40	589.5	16.8	720	10	09M2V7	09M2V7 arabidopsis
41	586.5	16.7	708	10	09M2V5	09M2V5 arabidopsis
42	582	16.6	1328	5	09NGP5	09NGP5 dictyostell
43	582	16.6	1509	5	08T688	08T688 dictyostell
44	579.5	16.5	627	11	091WA9	091WA9 mus musculu
45	579.5	16.5	646	11	08K4E1	08K4E1 mus musculu

ALIGNMENTS

RESULT 1

ID 08R543 PRELIMINARY; PRT; 673 AA.

AC 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE Sterolin 2.

GN ABCG8

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_Taxid=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=129/SV;

RA Lu K., Zhou Y., Lee M.-H., Patel S.B.;

RT "Molecular cloning, genomic structure and characterization of novel

mouse head-to-head tandem ABC transporters.";

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF351811; AL82898.1; JOINED.

DR EMBL: AF351800; AL82898.1; JOINED.

DR EMBL: AF351801; AL82898.1; JOINED.

DR EMBL: AF351802; AL82898.1; JOINED.

DR EMBL: AF351803; AL82898.1; JOINED.

DR EMBL: AF351804; AL82898.1; JOINED.

DR EMBL: AF351805; AL82898.1; JOINED.

DR EMBL: AF351807; AL82898.1; JOINED.

DR EMBL: AF351808; AL82898.1; JOINED.

DR EMBL: AF351809; AL82898.1; JOINED.

DR EMBL: AF351810; AL82898.1; JOINED.

DR InterPro: IPR004439; ABC_transporter.

DR Pfam: PF00005; ABC_tran; 1.

DR ProDom: PD000006; ABC_transporter; 1.

DR PROSITE: PS00211; ABC_TRANSPORTER; 1.

SO SEQUENCE 673 AA; 7608 MW; FA08340445DF259C CRC64;

Query Match 81.9%; Score 2871; DB 11; Length 673;

Best Local Similarity 81.8%; Pred. No. 3,5e-216;
Matches 551; Conservative 52; Mismatches 69; Indels 2; Gaps 2;

QY 1 MAGKAAEERGLPGKATPDOTS-GLQDLRFSSSDNSLYFTYSGQPNLTVLRDLYOYDLA 59
DB 1 MAEKTEETQLMNGTVLQDASQGLQDLRFSSSDNSLYFTYSGQSNLTVLRDLYOYDLA 60
QY 60 SOVPWFQDLAQFKPMPTSPSCQNSCELGIONLSFKRVSGOMLAIIGSSGGRASLDVIT 119
DB 61 SOVPWFQDLAQFKPMPTSPSCQNSCELGIONLSFKRVSGOMLAIIGSSGGRASLDVIT 120
QY 120 GRHGGRKSGQIMNGOPSSPOLVRKCAVAVRHQNDLPLNLTVPRETLAIAQRRLPRTF 179
DB 121 GRHGGRKSGQIMNGOPSSPOLVRKCAVAVRHQNDLPLNLTVPRETLAIAQRRLPRTF 180
QY 180 SOAQRDRVEDVIAELRLQCAQTRVGNMVRGLSGGERRRVSIGVOLLNPGILLIDEP 239
DB 181 SOAQRDRVEDVIAELRLQCAQTRVGNMVRGLSGGERRRVSIGVOLLNPGILLIDEP 240
QY 240 TSGDSTFANHLVYTLRLAKGNRLVLSLHQPRSDIFRLFDVLLMTSGTPYIYGAQH 299
DB 241 TSGDSTFANHLVYTLRLAKGNRLVLSLHQPRSDIFRLFDVLLMTSGTPYIYGAQH 300
QY 300 MVOYFAGYPCPRYSNPADFYVDLTSIDRSRQELATREKQSLAALFLKVRDLDDF 359
DB 301 MVOYFAGYPCPRYSNPADFYVDLTSIDRSRQELATREKQSLAALFLKVRDLDDF 360
QY 360 LMKAEKRLDDEDCVSSVTPPLDNTCLPSPFKMGAVQOFTTLIRROISNDPDLPTLLH 419
DB 361 LMKAEKRLDDEDCVSSVTPPLDNTCLPSPFKMGAVQOFTTLIRROISNDPDLPTLLH 419
QY 420 HGEACLMSTIGFLYFGHSIQLSFMDTALLFMIGALIPFNVLIDVSKYSERAMLY 479
DB 420 HGEACLMSTIGFLYFGHSIQLSFMDTALLFMIGALIPFNVLIDVSKYSERAMLY 479
QY 480 YELEDGLYTGPFPAKILGELPERHCAYIIYGMPTVYLANLRPGLOPFLHLVWLYV 539
DB 480 YELEDGLYTGPFPAKILGELPERHCAYIIYGMPTVYLANLRPGLOPFLHLVWLYV 539
QY 540 FCCRIMALAALALPTFHMAFFSNALYNSFYLAGGFMINLSLMTVPAMISKVSFLRMC 599
DB 540 FCCRIMALAALALPTFHMAFFSNALYNSFYLAGGFMINLSLMTVPAMISKVSFLRMC 599
QY 600 FEGIMAKIOFSRRTYKMPGLNLTAVSGDKILSAMELDYPVLAIVLIGSGGFVLYV 659
DB 600 FEGIMAKIOFSRRTYKMPGLNLTAVSGDKILSAMELDYPVLAIVLIGSGGFVLYV 659
QY 660 VSLRFKOKPSODM 673
DB 660 VSLRFKOKPSODM 673

RESULT 2
OBCIO5 PRELIMINARY: PRT: 672 AA.

AC 08C105: 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
GN ABCG8.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN 11)
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RA Yu H., Lu K., Lee M., Pandit B., Patel S.B.;
RT "The rat Abcg5 and Abcg8: characterization, chromosomal assignment and
genetic variation in s10sterolemic rats.";
RT Submitted (Aug-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AY145899; AAN64276.1; -

SO SEQUENCE 672 AA; 75906 MW; 2FE0846E71BD9D47 CRC64;
Query Match 80.9%; Score 2835.5; DB 11; Length 672;
Best Local Similarity 79.9%; Pred. No. 2.1e-213;
Matches 538; Conservative 57; Mismatches 77; Indels 1; Gaps 1;

QY 1 MAGKAAEERGLPGKATPDOTSGLQDLRFSSSDNSLYFTYSGQPNLTVLRDLYOYDLA 60
DB 1 MAEKTEETQLMNGTVLQDASQGLQDLRFSSSDNSLYFTYSGQSNLTVLRDLYOYDLA 60
QY 61 OVPWFQDLAQFKPMPTSPSCQNSCELGIONLSFKRVSGOMLAIIGSSGGRASLDVIT 120
DB 61 OVPWFQDLAQFKPMPTSPSCQNSCELGIONLSFKRVSGOMLAIIGSSGGRASLDVIT 120
QY 121 GRHGGRKSGQIMNGOPSSPOLVRKCAVAVRHQNDLPLNLTVPRETLAIAQRRLPRTF 180
DB 121 GRHGGRKSGQIMNGOPSSPOLVRKCAVAVRHQNDLPLNLTVPRETLAIAQRRLPRTF 180
QY 181 SOAQRDRVEDVIAELRLQCAQTRVGNMVRGLSGGERRRVSIGVOLLNPGILLIDEP 240
DB 181 SOAQRDRVEDVIAELRLQCAQTRVGNMVRGLSGGERRRVSIGVOLLNPGILLIDEP 240
QY 241 TSGDSTFANHLVYTLRLAKGNRLVLSLHQPRSDIFRLFDVLLMTSGTPYIYGAQH 300
DB 241 TSGDSTFANHLVYTLRLAKGNRLVLSLHQPRSDIFRLFDVLLMTSGTPYIYGAQH 300
QY 301 MVOYFAGYPCPRYSNPADFYVDLTSIDRSRQELATREKQSLAALFLKVRDLDDF 360
DB 301 MVOYFAGYPCPRYSNPADFYVDLTSIDRSRQELATREKQSLAALFLKVRDLDDF 360
QY 361 LMKAEKRLDDEDCVSSVTPPLDNTCLPSPFKMGAVQOFTTLIRROISNDPDLPTLLH 420
DB 361 LMKAEKRLDDEDCVSSVTPPLDNTCLPSPFKMGAVQOFTTLIRROISNDPDLPTLLH 420
QY 421 GAEACLMSTIGFLYFGHSIQLSFMDTALLFMIGALIPFNVLIDVSKYSERAMLY 480
DB 421 GAEACLMSTIGFLYFGHSIQLSFMDTALLFMIGALIPFNVLIDVSKYSERAMLY 480
QY 481 YELEDGLYTGPFPAKILGELPERHCAYIIYGMPTVYLANLRPGLOPFLHLVWLYV 540
DB 481 YELEDGLYTGPFPAKILGELPERHCAYIIYGMPTVYLANLRPGLOPFLHLVWLYV 540
QY 541 CCRIMALAALALPTFHMAFFSNALYNSFYLAGGFMINLSLMTVPAMISKVSFLRMC 600
DB 541 CCRIMALAALALPTFHMAFFSNALYNSFYLAGGFMINLSLMTVPAMISKVSFLRMC 600
QY 601 FEGIMAKIOFSRRTYKMPGLNLTAVSGDKILSAMELDYPVLAIVLIGSGGFVLYV 660
DB 601 FEGIMAKIOFSRRTYKMPGLNLTAVSGDKILSAMELDYPVLAIVLIGSGGFVLYV 660
QY 661 VSLRFKOKPSODM 673
DB 661 VSLRFKOKPSODM 673

RESULT 3
O9ARU4 PRELIMINARY: PRT: 668 AA.

AC 09ARU4: 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
GN P0445D12.3.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OX Euphorbiaceae; Oryzae; Oryza.
OX NCBI_TaxID=4530;
RN 11)
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsunoto T., Yamamoto K.;

Oryza sativa n1pnonbare(GA3) genomic DNA, chromosome 1, PAC clone: P0445D12.
 RT Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL: AP003046; BAB40032.1; -
 DR Gramene: O9AR04; -
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transporter; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 DR ATP-binding; Transport.
 KM Sequence: 668 AA: 73368 MW: D1875B8C75B0F3B2 CRC64;
 SQ
 Query Match 21.2%; Score 742; DB 10; Length 668;
 Best Local Similarity 30.3%; Pred. No. 1.8e-49;
 Matches 188; Conservative 124; Mismatches 252; Indels 56; Gaps 12;
 QY 75 WTSPSCNSCELG-----IQNLSPYVSQGMALIGSSGGRASLDVITGR--GHGK 126
 Db 58 MARITCALKNRGDVARFLSLNSGKSKRLALMGPSGKTTLLNVLAGULTASPL 117
 QY 127 IKSQIWINQSPSSPOLVRKCAVHROHQLPNTLVRETIAFIAQMRTPRPSQAORDK 186
 Db 118 HLSGFLYINGRPISSEGYK--IAYVROEDLFESQLVRETLSIAELQRLTLPERKES 175
 QY 187 RVEDVIAELRLRQCADTRVGNMNVYRLSGGERRRVSIQVOLLNPGILLDEPTSGIDST 246
 Db 176 YVNDLFLRGLVNCADISVDKARVKGISGGERKRLSLACELASPSIITFDEPTGIDAF 235
 QY 247 TANNLVKTLRLKAGNRLVLSLHOPRSDIFRLFDVLMTSGTPIYLG-AAQHMVQYF 305
 Db 236 QAEKVMETLRQLAEDGHTVICSIHPRGSYVKGFDIVLSBEVYIMKAPKEPLLYRA 295
 QY 306 AIGYPPRYSNRPFVYDLTSDRSREDELATREKAOSIALFLERKVDLDLFLKAEV 365
 Db 296 SLGYHCPDHVNPALFLADLISVDSAESVOSRKRILENIEEFSNKV-----AT 346
 QY 366 KDLDEDTCESSVT-PLDINCLP-----SPTK-MRPAVOFTLLIRQISNDFRDLPTLL 418
 Db 347 ES-----NSSLTNDEGSEFSPSKLIQKSTTKHRRGMWRQFLFRAMQOAFRDGPYTK 399
 QY 419 IHGAECALMSMTIGLFGHGSITOLSPMDTALLFPGALIPFNVLIDVISCYSERAML 478
 Db 400 VARBSVASAIIFGSYFWMKGTQTSIODRMGLIQTALINTAMALTKTVGCFPKERRAIV 459
 QY 479 YLEDEGLYTTGPFYFAKILGELPEHCAYIITYGMPTWLANLRPGIQLPFLHFLVWLV 538
 Db 460 DBERAKGYALGPYSSKLLAEPIGAAPLIFGSLIYPMKLPFFSFAFCGIVTAE 519
 QY 539 VFCRCRIMALAAALPTFHMASFNSALYNSFYLAGFPMINLSLMTVAAMSKYSFLM 598
 Db 520 SFRASAMGLTGMATTEAAMALGPSLMTVFIVFGYVNPDMNPVIRWIMPKVSLIMW 579
 QY 599 CEEGLAKIQF-----SRTYKMPGLNLTIAVSGDKLISAMELDSPLAIIYVIGLSG 652
 Db 580 AFGGLCINFEKGLQEPQOHSYDIOGE-----QALNRSGLGIRIADTLVAO 626
 QY 653 GEMVLYVSLRFI---KQK 669
 Db 627 GRILMFVWLYTLTKKNRP 646
 RESULT 4
 O9L182. PRELIMINARY; PRT; 672 AA.
 AC O9L182;
 DT 01-OCT-2000 (TRENBLREL. 15, Created)
 DT 01-OCT-2000 (TRENBLREL. 15, Last sequence update)
 DE 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
 DE ABC transporter-like protein.
 OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
 OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;
 OC eucosids II: Brassicales: Brassicaceae: Arabidopsis.
 OC NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RX MEDLINE=20363099; PubMed=10907853;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety pl.
 RT TAC and BAC clones.";
 RL DNA Res. 7:217-221(2000).
 DR EMBL: AP001313; BAB03081.1; -
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transporter; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 DR PROSITE: PS00012; PHOSPHOPANTHEINE; 1.
 DR ATP-binding;
 KM Sequence: 672 AA: 75269 MW: 20B2D99215600135 CRC64;
 SQ
 Query Match 21.1%; Score 739.5; DB 10; Length 672;
 Best Local Similarity 30.6%; Pred. No. 2.8e-49;
 Matches 221; Conservative 126; Mismatches 251; Indels 125; Gaps 27;
 QY 7 EERGLPK--GATPQDTSGLDRLFSSES-----DN-----SLFTYSQGN 45
 Db 7 QESSFPKTSANHHERSPVOENRFSFHYNPCLDDNDHDSHRSQSSVLRSLSR 66
 QY 46 TLEVRDLNTQVD-----LASQVWEQLAQFMPTSPSCNSCELGIONLSFKVR 96
 Db 67 ILKFEELFYISKQSGKSGSYWFGSQEPKRNRLV-----KCVSGI-----VK 108
 QY 97 SQGMALIISSGGRASLDVITGRGHGKIKSQIWINQSPSSPOLVRKCAVHROHNO 156
 Db 109 PGLLAMLPLPSGSGKTLVTALAGRLQ-GKL-SGVSYNGEPTSSVKRK-TGFVYQDV 165
 QY 157 LPLNLTVRETIAFIAQMRTPRPSQAORDKRVEDVIAELRLRQCADTRVGNMNVRLSSG 216
 Db 166 LYPHLVMTLVYTLALRLPKELTRKELEOVEMVYSDGLTRCCNSVIGGLIRKISGG 225
 QY 217 ERRRVSIGVOLLNPGILLDEPTSGIDSTFTHNLVKTLSRLAKNRLVLSLHOPRSDI 276
 Db 226 ERKRVISIGEMLVNPELILLDEPTSGIDSTTAARIATVRLSLARGRTVVTIHQSSSL 285
 QY 277 FRLLFDVILMTSGTPIYLAQAQHMVQYFAIGY-PCPRYSNRPADFVDT---SIDRSR 332
 Db 286 YRFEDKVLVLSGCPYISGDSGRVMEFGSIGVOPSSVFNPADEVLDIANGITSDTKY 345
 QY 333 EQLATREKAOSIALFLERKVDLDLFLKAEVKDLDEDTCESSVT-PLD----- 382
 Db 346 DO-IETNGILDR-----LEQNSVKOSLISSTKKNLYPPLKEEVSTFEPQDPTNARLRK 399
 QY 383 --TNCLPSPTKMPGAVQFTLLIRQIS-----NDFRDLPTLLIHGAECALMSMTIGLYE 436
 Db 400 AITNRMPISMM-----QSVLLKNGKLRSHESFGLRIFFVMS-----VSLSGILMW 449
 QY 437 GHGSIQSLSPMDTALLFPGALIPFNVLIDVISCYSERAMLYLEDEGLYTTGPFYFAK 496
 Db 450 -HSRV-AHLQDQVGLFFPSIFWGFPELNAIFTEPOERPMILIKRRSGIYRLSYIAR 507
 QY 497 ILGELPEHCAYIITYGMPTWLANLRPGIQLPFLHFLVWLVVFCRCRIMALAAALPTF 556
 Db 508 TVGDLPMEILPTIFVTITVMGKPSLFTFIMTLMVLYLVNVAQVGLAIGAILMDA 567

QY	243	LDSEFAHNLVVTLSRLKAGNRLVLTSLHQPSPDIFRLVDLVLMTSGPTIYG-AAQHNV	301
Db	246	LDAPFAEKVETKIAKQDGHFTVLCISIHQPGSVYAKFEDVLVLTBGLVYAGAGKEPL	305
QY	302	QYFAIATGPCRYSNPADEYVDLTSIDRSREDELATREKAQSIATLAFLEKVRDLDFLM	361
Db	306	TYFGNFGFLCEHNVAPEFLADLLTSVDYSSEFEYVSQKRVHALVDAFSGR-----	356
QY	362	KAETKDLDEDICVSSV---TPLDNCLPSTK-----NPGAVQETTLIR	405
Db	357	-----SSSVLYATPLS---MKKEETKNGMRRRAIYERTDGMWRQFFLLKR	400
QY	406	QISNDFRDLPTLLIHGEACIMSMTIGFLYGHGSIOLSPFMDTAALFMICALPFWYL	465
Db	401	AMMAQSRDGPINXKVRARMSVASANVIFGSEFMRMKSQTSIODRGLL-QVAAT---NTAM	456
QY	466	DVISCY-----SEPAMLYLEEDGLYTTGPYFAKILGELPEHCAYIIITYGMPYYMLNT	521
Db	457	AALTKYGVCFPEKERAIVDRERSKGSYSIGPYLLSKTIAEIPIGAAPLMFGAVLYPYARL	516
QY	522	RPGLQPLFLHRLVWLIVVCFCCRIWALAAALPPEFHMASPFSSNMLYSPFIAGFMTNLS	581
Db	517	NPTLSRGKFCGCIYTVSEFSAASAGLTVGAVWPSTEAMAVGSPMLTVFVFGYYVAD	576
QY	582	SLMTVPAMWISKVSLRMCFEQIMKIOES-----RRTYKMLPGLNT---IAVSGDKTISA	632
Db	577	NTPILFPHWIPASLIIRMAFGCLCINFSGLKFPDQNTFVDVOTGEQALERLSFGSRRIRET	636
QY	633	MELDSYPL-----ATYLIIV	647
Db	637	IAAOSRIIMFWYSATYLLL	655

RESULT 6

Q9ASR9.

ID Q9ASR9 PRELIMINARY; PRT: 725 AA.

AC Q9ASR9;

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE ACg01320/F10A8.20.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eudicotyledon: Viridiplantae: Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae; OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.

OC NCBI_Taxid=3702;

OR

RN

RP SEQUENCE FROM N.A.

RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J.,

RA Bowser L., Carninci P., Chung M.K., Goldsmith A.D., Hayashizaki Y.,

RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,

RA Lee J.M., Lin Y., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,

RA Palm C.J., Toriumi H., Sakurai T., Satou M., Seki M., Southwick A.,

RA Shinozaki K., Davis R.W., Theologis A., Toriumi M., Yamada K., Yu G., Yu S.,

RT "Arabidopsis ORF clones."

RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

RN

RP SEQUENCE FROM N.A.

RA Cheuk R., Chen H., Kim C.J., Shinn P., Banh J., Bowser L.,

RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,

RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,

RA Lee J.M., Lin Y., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,

RA Palm C.J., Toriumi H., Sakurai T., Satou M., Seki M., Southwick A.,

RA Shinozaki K., Davis R.W., Theologis A., Toriumi M., Yamada K., Yu G., Yu S.,

RT "Arabidopsis ORF clones."

RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

CC -i SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.

DR EMBL: AF367318; AAK32905.1; -

DR EMBL: AY133617; AAM91447.1; -

DR InterPro: IPR003439: ABC transporter.


```

OY 307 IGYPCRYSNPADFYVDLTS-----IDRSRBEOLATREKASLALFLEKVRDLDDEFLM 361
DB 298 LQFSTSLTVNPDLLDLNGLIPPTOKETSEOEKTVK--ETLVSAYEKNI----- 347
OY 362 KAEKFDLDETCVESS-----VPLDTNCLPSPKMGAVOQFTTLIRROI-SNDRDLPT 416
DB 348 --STR-LKAEICNAMESHYETKAANKLSEQKCTTMYOTVLQGRRRRRESFNK 404
OY 417 LLIHGAECALMSMTIGLYFGHGSIQLSFMDTALLFMIGALIPFNVLIDVSKYSERA 476
DB 405 LRIF---OVISAFVGLLMMH-TPKSHIQDRTALLFFSVWGWGYPLYNVFTPOEKR 460
OY 477 MLYVELDEGLTGTGYFFPAKILGELPEHCAYIIIGMPTWLANRPGLOPFLHPLW 536
DB 461 MLIERSSGMRYLSSSYFARVNGDLPLELALPTAFVFTIYMWGIGLPPDTEFLSLVL 520
OY 537 LVFECRCRIMALAAALLPTFHMAFSPSNALYNSFYLAGFMNLSLMTVP---AMISKV 593
DB 521 KSVLVAOGLGALFGLMLNINIKATTLASVTLVFLIAGGYVO-----QIPFTYWLKYL 575
OY 594 SFLMCFEGLKMKIOPSRRTY-----KMPGLNLTIVASGDKILSAMEL 635
DB 576 SYSYCYKLLIGIOTDDDYECSSKWCRCVGDPEPAIKSMGLNNLMI----DVFVWGYVL 631
OY 636 DSYPLATYLLVIGLSGCFWVLYYSLR 663
DB 632 VGYRLMA-----YMALHRYKLR 648

RESULT 8
OY 09C6R7 PRELIMINARY; PRT; 646 AA.
AC 09C6R7;
DB 01-JUN-2001 (TREMBLrel. 17, Created)
DB 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DB 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DB ABC transporter, putative.
GN F5M6.22.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC OC eudicots II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Becker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujil C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Miltischer J., Miranda M., Nguyen M., Nieman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Tortum M.J., Town C.D.,
RA Uteberck T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT *Sequence and analysis of chromosome 1 of the plant Arabidopsis
R7 thaliana*.
RL Nature 408:816-820(2000).
DR EMBL, AC079041; AAG50724.1.
DR InterPro, IPR003593; AAA_Atpase.
DR InterPro, IPR003439; ABC_Transporter.
DR Pfam, PF00005; ABC_tran; 1.
DR ProDom, PD000006; ABC_Transporter; 1.
DR SMART, SMO0382; AAA; 1.

```

```

DR PROSITE, PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding.
SQ SEQUENCE 646 AA; 72342 MW; 7A9624F82DE88A6E CRC64;
Query Match 20.6%; Score 723.5; DB 10; Length 646;
Best Local Similarity 30.6%; Pred. No. 4,8e-48;
Matches 208; Conservative 119; Mismatches 262; Indels 91; Gaps 20;

```

```

OY 22 GLQDRLFSSSDNSLYF--TYSQPN-----TLEVRDLYOVDLASQVPMFQLOAFKMP 74
DB 20 GLPD-MSDTQKSVLAFPTTISQPOLQSMHPTILKEVYKVT-----EOTSQCKMS 71
OY 75 WTSFSCNSCELGIONLSFKYRSQGMALIGSSGGRASLDVITRGHGKIKSGQIWI 134
DB 72 WSKSE-----KTLINGITGVNCPGFEFLAMLGSSGSKTLLLSALGR--LSFTFGKVMY 124
OY 135 NGOPSSPOLVKCAVAVQHOLPLNLYVRELIATIAORLPRTFSQARQKVEDVIAE 194
DB 125 NGOPSSGCIKRR-TGEVADVDVLYPHLTVWETLFTFALLRLPSSLTREKEHVDVIAE 183
OY 195 LRLROCADTRVGNMVRGLSGGERRRVSIQVOLLNPGILLIDPTSGLDSEFTAHNLVKT 254
DB 184 LGINACTSMIGLPLFRGISGEEKRVSIGOBMLNPSLILDEFTSGLDSTTAHRIYTT 243
OY 255 LSLRAGNRRLVLSLHOPRSDIFRLFDVLLMTSGTPYTLGAAQHMVQFYTAIGPCRY 314
DB 244 IKRLASGGRTVYVTTIHOPSSRIYHMFEDVLLSESPYLYGAASAVERYFSSLSGSTSLT 303
OY 315 SNPADFYVDLTS-----IDRSRBEOLATREKASLALFLEKVRDLDDEFLMKAETKOLD 369
DB 304 VNPADLLDLNGLIPPTOKETSEOEKTVK--ETLVSAYEKNI-----STK-LK 350
OY 370 EDTCVESSE-----VPLDTNCLPSPKMGAVOQFTTLIRROI-SNDRDLPTLIHGMEA 424
DB 351 AELCAHESHYETKAANKLSEQKCTTMYOTVLQGRRRRRESFNKLRP---Q 407
OY 425 CIMSMTIGLYFGHGSIQLSFMDTALLFMIGALIPFNVLIDVSKYSERAMLYELED 484
DB 408 VISVAFGLGLMMH-TPKSHIQDRTALLFFSVWGWGYPLYNVFTPOEKRMLKKESS 466
OY 485 GLYTTGPEFFPAKILGELPEHCAYIIIGMPTWLANRPGLOPFLHPLWLVVFCRI 544
DB 467 GMTRLSYFARVNGDLPLELALPTAFVFTIYMWGIGLPPDTEFLSLVLYSVLVAG 526
OY 545 MALAAALLPTFHMAFSPSNALYNSFYLAGFMNLSLMTVP---AMISKVFLRWCE 601
DB 527 LGLAFGALMLNINIKATTLASVTLVFLIAGGYVO-----QIPFTYWLKLYSYCYK 581
OY 602 CLMKIOPESRRTY-----KMPGLNLTIVASGDKILSAMELDSYPLVYAI 643
DB 582 LLLGIQYDDDYEECSKWCRCVGDPEPAIKSMGLNNLMI----DVFVWGYVLVGYRLMA- 636
OY 644 YLYIYIGLSGCFWVLYYSLR 663
DB 637 -----YMALHRYKLR 646

RESULT 9
OY 08C104 PRELIMINARY; PRT; 652 AA.
AC 08C104;
DB 01-MAR-2003 (TREMBLrel. 23, Created)
DB 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DB 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DB Stereolign 1.
GN ABCG5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;

```

RA Yu H., Lu K., Lee M., Pandit B., Patel S.B.:
 RT "The rat Abog5 and Abog8: characterization, chromosomal assignment and
 genetic variation in six tolemaic rats."
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY145599; AAN64275.1;
 SQ SEQUENCE 652 AA; 73372 MW; 49FE7372269299D CRC64;

Query Match 20.38; Score 713; DB 11; Length 652;
 Best Local Similarity 30.08; Pred. No. 3.2e-47;
 Matches 190; Conservative 115; Mismatches 232; Indels 96; Gaps 15;

QY 12 PKGAT-PODTSGLDRLFFSESDNSLYFTYSGQNTLEVDLYVDLASOV-PWEQLA 69
 Db 9 PEARQPHNRGSO-----SLEKGSV--TSEARHSGV--LWFSVSNRVGPW----- 55
 QY 70 QFMPTSPSCQSCELGI-QNLSFKVSGOMLAIIGSSCGRASLIDVTGSGHGKIK 128
 Db 56 -----WNKSCQCKMKRKLKDVSLYESQTCILGSSSGKTTLLDLSGLRRTGTL 110
 QY 129 SGQIWINGOPSSPOLYRKCAVAVRHQNLPLNLTVEBTLAFIAOMRLPRTSQORDKRY 188
 Db 111 EGEVFNCGELRRDQDFCVSYLLQSDVFLSLTVEETLKYTMAL- RSSADPYDKV 169
 QY 189 EDVIAELRQCADTVGNMAYVGLSGERRRISIGVQLMNPGLILDEPTSGDLSFTA 248
 Db 170 EAVTELSSHVADOMIGNNNEGSGERRRISIAOILDQPKVMMLDEPTGDCMTA 229
 QY 249 HNLVKTLSRLAKNRVLTLSDHOPRSDFRLPDLVLTMTSGPIYVIGAOMHVOYTAIG 308
 Db 230 NHIVLLVLLARLAKRIVITIHOPRSELHHEPKIALITYGELVFCGTEPMGLFNNCG 289
 QY 309 YCPKPSNPADYVDLTIDRSRQELATREKAOIALF-----LEKVRDL 356
 Db 290 YCPKPSNPADYVDLTIDRSRQELATREKAOIALF-----LEKVRDL 356
 QY 357 DDFWKAETKDLDEDCVSSVTPDLNCLPSTK-MPGAVOFTLLIRQISNDRDLP 415
 Db 350 -----KTLPM-----VPEFTKMPGMECKLGVLLRRVRLNLRNKO 385
 QY 416 TLLIHGAELMSMTIGF-LYFGHSIQLSFDPTAALFMIGALIPENVIDVSKCS 473
 Db 386 VYIMLYOQLMIGLIFLTLVONNMMLGAVDRLGYLVGATPYGMNAVALFPM 445
 QY 474 EPMALYELEDGLYTPYPAKILGELPEHCAYIIYGMPTWLANLRGLOPFLHFL 533
 Db 446 LRAVDSQESODGLYOKMQLAYVHALPFSIVATVIFSSVYKWLGLYEVARF----- 500
 QY 534 LVMVLVFCRIMAAALLPFFHMASPFSNAL-----YNSFYLAGG 575
 Db 501 -----GYFSALMLAPHLIGELFTLVLMGNONPINSIVALLISGLIGSG 548
 QY 576 FMINLSLMTVPAMISKVSFLRCEGLMKIOE 608
 Db 549 FIRMIEPMPIKILGTFPOKTCCEILVNEF 581

RESULT 10

ID 0949Y4 PRELIMINARY: PRT: 662 AA.
 AC 0949Y4:
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Putative ABC transporter protein.
 GN F17M19.11.
 OS Arabidopsis thaliana (mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.-X., Pham P.K., Banh J., Banno F., Dale J.M.,

RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B.,
 RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Satou M., Seki M., Shim P., Southwick A., Tracy S.E.,
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.:
 RT "Full length cDNA of gene F17M19.11 (GI:13324545)."
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY50810; AAK92745.1;
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transporter; 1.
 DR SMART: SM00382; AAA; 1.
 KW ATP-binding.
 SQ SEQUENCE 662 AA; 72903 MW; CD5BC0853261BC45 CRC64;

Query Match

Best Local Similarity 20.28; Score 709; DB 10; Length 662;
 Matches 216; Conservative 107; Mismatches 233; Indels 136; Gaps 23;

QY 44 PNTLEVRDNLTYQVDLASOVWFEQLAQFKMPTSPSCQSCELGI-QNLSFKVRS----- 97
 Db 37 PTLKRFVDCYRKIKGM-----SNDSCNIKLLGLKOKSDETRSTBERT 82
 QY 98 -----GOMLAIITSSCGGRASLIDVTGGRGHGKIKSGQIWINGOPSSPOLYRK 146
 Db 83 ILISGYTGMTISPGFEMVILGPSGSGKSTLLNVAVGRHGSNL-TGKILINDGKTKTQIKR 141
 QY 147 CYAVRHQNLPLNLTVEBTLAFIAOMRLPRTSQORDKRYEDVIAELRQCADTVRNG 206
 Db 142 -GFTVADQLPLHLVRETLVALLRPRSLTRQYKRAESVISELGTICENTVVG 200
 QY 207 NMVRLSGGERRRISIGVQLMNPGLILDEPTSGDLSFTAHLNKTLSRLAKG-NRLY 265
 Db 201 NTFIRISGGERKRRVSIABELLNPSLVLDERTSGDLTAALRQTLGALHAGKTV 260
 QY 266 LISLHOPRSDFRLPDLVLTMTSGPIYVIGAOMHVOYTAIGPCPRSNADYVDLT 325
 Db 261 VTSIHOPSSSRVQMEFTVLLSEGKCLFVGKGRDAMAYESGFSAPFANPADFLDLA 320
 QY 326 SIDRSRQELATREK-----AOSIALFEKVRDLDFLMAKETKDLDEDCVSSVTPDL 382
 Db 321 --NGVCQDDGVTEREPNVRQTLVYAY-----DTLLAPQVK-----TCIEVSHPD 365
 QY 383 TNCLEPSTKMPGAVOFTLLI-----RQISNDRDLP.TLLIHGAELC 425
 Db 366 -NARVKTFRVNGG--GITTCIATWFSQCLILHLRLLKERHESD-----LRLFOV 415
 QY 426 LMSMTIGLYFGHSIQLSFDPTAALFMI-----GALIPFNVLIDVSKYSERAMLYE 481
 Db 416 AASILGLMMW-HSDYR-DVHDLGLIFLISIFWGLVPSFNVPF-----PERRAFTTE 469
 QY 482 LEDGLYTPYPAKILGELPEHCAYIIYGMPTWLANLRGLOPFLHFLVWLVC 541
 Db 470 RASGMYTSSYFMAVLGSLSMELVLPASFLFTFYMWVYLARGIYVFLTLTLVLLYVLA 529
 QY 542 CRIMLAAALLPFFHMASPFSNALNSFYLAGGFMINSLSMTVPA--WTSKSYFLMW 598
 Db 530 SOGLGLAIGAMIMAKKASTIYTWMLAFVLTGTYVN-----KVPSGVMWKVYSTFY 584
 QY 599 CFEGLMKIQFSRRYKMPDGLNLTIVASGDKTILSAELDYSPLYA-----IYLLIV 647
 Db 585 CYRLVLAIOYG-----SGEILIRMLGCGSKGOGASNAATSGACRVEEYV 629
 QY 648 IGLSG-----GFWLYVYSIRFK 666
 Db 630 IGDVGMTSVGLFLMFGYRLVLAIALRIK 661

RESULT 11

Q9C8W6

ID 09C8W6 PRELIMINARY: PRT: 609 AA.
AC 09C8W6;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Putative ABC transporter.
GN F17M19.11.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustoids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Elgu P., Feldblum T.V., Feng J.-D., Feng B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maiti R., Marziani A.,
RA Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterback T., Van Aken S., Vayenberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana."
RT Nature 408:816-820(2000).
RL EMBL: AC021665; AAG52231.1; -
DR InterPro: IPR003593; AAA_Atpase.
DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF00005; ABC_tran.1.
DR Prodom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
KW ATP-binding.
SQ
SEQUENCE 609 AA; 67007 MW; 65D11A874E5C0B61 CRC64;
Query Match 20.0%; Score 700; DB 10; Length 609;
Best Local Similarity 31.8%; Pred. No. 3,1e-46;
Matches 210; Conservative 103; Mismatches 225; Indels 122; Gaps 22;
OY 76 TSPSCNSCELG-I-ONLSFKVRS-----GOMLAIIGSGCGRASLDVI 118
DB 2 SNDSCNIRKLGLKPKPSDEFRTSTERTILSCVTGMISGERMAVLGPSSGSKTLNANV 61
OY 119 TGRGSGRIKSGQIMINGOPSSPOLYRKCVANVRHNOILPMLVIRETLAFLAQRDPT 178
DB 62 AGRHLSGML-TKILINCKITKQTLKR-TGFVADDDLLPRLVRETLVFAALLRPRS 119
OY 179 FSOAQRDRVEDVIAELRLQCADTRVGMVMYRGSGERRRVSIGVOLLNPGITLIDE 238
DB 120 LTRVOKLAARESVISELGLTKCENTVNGTFRIGSGGERKRVSLAHELINPSLVIDE 179
OY 239 PTSGIDSTFANLVKTLISRLAKG-NRLVLSLAHPRSDIFRLDLVLLMTSGTPIYLGA 297
DB 180 PTFSGIDATAALRLVOTLAGLHGKGTAVTSIHOPSSHVQFOWFDVLLLSGKCLFVGK 239
OY 298 OHMVOYFAIGPCPRYSNPADFYDLTISDRSRSEGLATREK---AQLAALFLEVR 354
DB 240 RDAKAYFESVSGSPAFPMNPAFDLIDL--NGVCTQDVTGEKRNVRQTLVTAY----- 292
OY 355 DLDDFLMAETKRLDEDTCESSVYPLDTNCLPSTKMPAGVVOGFTLLI----- 403
DB 293 ---DTLAPQVK-----TCIEVSHPPQD-NARFVKTIRNG--GITTQIAIWFQGLCIL 341
OY 404 -----RROISNDPFDLFTLLIHGAELACMSMTIGFTLVFGHGSIOLSFMDPALILMI-- 455

DB 342 HRLKERRHESFD-----LIRIQVNAASILCGLMM-HSDYR-DVADRGLGLFFISI 392
OY 456 --GALIPENVIIDVSKSCSEANMLYEEDLTYTTPYFFAKLIGELPEHCAYIITGM 513
DB 393 FNGVLPSEFNAVPTF-----POERAIFTRERASGMVTLSSYFMAHVAGLSIMELVLPASFLT 448
OY 514 PTYMANLRPGLOPFLHFLVWLVFCCRIMAAALLPFFHMASFFSNLINSFYLA 573
DB 449 FTYMAYLRPGVYPLFTLITLVLLVYASQIGLALGAALIMAKKASTIVVTMLAFVLT 508
OY 574 GGFMINLSLWTVPA---WISKVSEFLRMCFCGLMKIOFSRTYKKMPLNLTAVSGDKIL 630
DB 509 GGYYVN-----KVPSSGMVMKRYSTTFYCYRLVVAIQG-----SGEELI 548
OY 631 SAMELDSITPLVA-----ITLVIGLSG-----GPMVLYVSLERFIK 666
DB 549 RMIGDSDSKOGASATSAGCRFEVEEVYIGVGMVTSVGLFMEFGRVILAYLARIK 608
RESULT 12
ID 08T691 PRELIMINARY: PRT: 801 AA.
AC 08T691;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE ABC transporter ABCG1.
GN ABCG1.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Anjard C., Loomis W.F.;
RT "Evolution of the ABC transporters of Dictyostelium."
RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL: AF482380; AAL91485.1; -
DR InterPro: IPR003593; AAA_Atpase.
DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF00005; ABC_tran.1.
DR Prodom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transport.
SQ
SEQUENCE 801 AA; 90052 MW; CCC4F0036CB195A3 CRC64;
Query Match 19.8%; Score 635.5; DB 5; Length 801;
Best Local Similarity 27.9%; Pred. No. 1e-45;
Matches 187; Conservative 134; Mismatches 230; Indels 119; Gaps 20;
OY 88 IONLSFKVSGOMLAIIGSGCGRASLDVYTRGSGKIK-SGOIWINGOPSSPOLYRK 146
DB 139 LTNINGHIESGTIFRIMGSPGAGKTTLLDI--AHRININGSGIMYLNKSKSPENIKK 195
OY 147 CVANVRHNOILPMLVIRETLAFLAQMPLPTFSOARDKRVEDVIAELRLQCADTRVG 206
DB 196 LCGVYTOQSDSLMPSLIVRETLNFYAKLMPDVPLEKLRQVQDIIDEMGLRCADTLVG 255
OY 207 --NNYVRSLSGERRRVSIGVOLLNPNPILIDEPTSGDLSFTANLVKTLISRLAKGR 264
DB 256 TADNKIRISGGERKRVITSTELTGPVSVLLIDETPSGLDSTSYVSAALKLAKSGRT 315
OY 265 VLISLHPRSDIFRLDLVLLMTSGTPIYLGAOHMVOYFAIGPCPRYSNPADFYDL 324
DB 316 IICITHPRSNLYDMFDLWLLDGLGONTIYKNAKALEYFNANGYHCSEKTNPNPAFDL 375
OY 325 -----TSID----- 328
DB 376 INTOVEDQADSDDDVDNDEEBEIGGGGGSGGAGGIEDIGISISPTMGSAVDNINKN 435

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 17:14:49 ; Search time 55 Seconds

(Without alignments)
1453.188 Million cell updates/sec

Title: US-09-989-981a-8

Percent score: 3506
Sequence: 1 MAGKAERGLPKGATPQDT.....FMVLVYSLRFIKQPSQDW 673

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCR_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCRUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3506	100.0	673	11	US-09-989-981a-8
2	3502	99.9	673	15	US-10-090-455-7
3	2883.5	82.2	672	11	US-09-989-981a-4
4	697	19.9	672	9	US-09-837-992-3
5	697	19.9	651	11	US-09-989-981a-6
6	697	19.9	651	15	US-10-090-455-6
7	688.5	19.6	652	9	US-09-837-992-1
8	688.5	19.6	652	11	US-09-989-981a-2
9	666	19.0	657	10	US-09-866-866a-14
10	656	18.7	663	14	US-10-108-605-245
11	642.5	18.3	655	10	US-09-981-353-35
12	640.5	18.3	655	15	US-10-120-687-61
13	640.5	18.2	655	11	US-09-961-086-1
14	638.5	18.2	655	10	US-09-866-866a-10
15	638.5	18.2	655	15	US-10-090-455-5

16	634.5	18.1	655	10	US-09-866-866a-27	Sequence 27, Appl
17	618	17.6	638	14	US-10-072-621-10	Sequence 10, Appl
18	617	17.6	674	15	US-10-090-455-4	Sequence 4, Appl
19	609.5	17.4	545	15	US-10-083-357-1335	Sequence 1335, Ap
20	573.5	16.4	627	15	US-10-090-455-8	Sequence 8, Appl
21	572.5	16.3	646	14	US-10-154-452-4	Sequence 4, Appl
22	571.5	16.3	646	15	US-10-090-455-13	Sequence 13, Appl
23	570.5	16.3	646	14	US-10-154-452-8	Sequence 8, Appl
24	567.5	16.2	646	15	US-10-072-621-9	Sequence 9, Appl
25	567.5	16.2	646	15	US-10-090-455-2	Sequence 2, Appl
26	518	14.8	604	9	US-09-745-763-197	Sequence 197, App
27	464	13.2	1564	10	US-09-801-368-244	Sequence 244, App
28	397	11.3	1501	10	US-09-801-368-346	Sequence 346, App
29	385	11.0	1511	10	US-09-801-368-250	Sequence 250, App
30	359	10.2	780	15	US-10-156-761-8698	Sequence 8698, Ap
31	336	9.6	843	15	US-10-156-761-13999	Sequence 13999, A
32	252.5	7.2	894	9	US-09-741-669-434	Sequence 434, App
33	251.5	7.2	1543	15	US-10-005-3388-8	Sequence 8, Appl
34	250.5	7.1	1617	15	US-10-005-3388-6	Sequence 6, Appl
35	246.5	7.0	331	9	US-09-815-242-11909	Sequence 11909, A
36	243.5	6.9	1617	14	US-10-090-453a-2	Sequence 2, Appl
37	238.5	6.8	275	15	US-10-156-761-12256	Sequence 12256, A
38	238	6.8	352	9	US-09-815-242-10053	Sequence 10053, A
39	237.5	6.8	1594	10	US-09-971-121-4	Sequence 4, Appl
40	237.5	6.8	1642	10	US-09-971-121-2	Sequence 2, Appl
41	237	6.8	388	15	US-10-156-761-10969	Sequence 10969, A
42	235.5	6.7	1638	14	US-10-090-458-2	Sequence 2, Appl
43	235.5	6.7	1642	14	US-10-090-458-5	Sequence 5, Appl
44	235.5	6.7	1642	15	US-10-005-3388-5	Sequence 5, Appl
45	231.5	6.6	203	15	US-10-252-819-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-09-989-981a-8
Sequence 8, Application US/0998981a
Publication No. US20030049730A1
GENERAL INFORMATION:
APPLICANT: Hobbs, Helen H.
APPLICANT: Shan, Bel
APPLICANT: Barnes, Robert
APPLICANT: Tian, Hui
APPLICANT: Tularik Inc.
APPLICANT: Board of Regents, The University of Texas System
TITLE OF INVENTION: ABC5 and ABC8: Compositions and Methods of Use
FILE REFERENCE: 018781-007320US
CURRENT APPLICATION NUMBER: US/09/989, 981a
PRIOR FILING DATE: 2002-07-23
PRIOR APPLICATION NUMBER: US 60/252, 235
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/253, 645
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 8
LENGTH: 673
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: human ABC8 (hABC8)
US-09-989-981a-8
Query Match 100.0%; Score 3506; DB 11; Length 673;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 673; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAGKAERGLPKGATPQDTSGIQRLEFSESSNSLYFTYSGQNTLEVDLWYQDILAS 60
|||||
Db 1 MAGKAERGLPKGATPQDTSGIQRLEFSESSNSLYFTYSGQNTLEVDLWYQDILAS 60
QY 61 QVWFQDLAQFKPMWTPSCQNSCELGIONLSFKVRSQMLALIGSGGCRASILDVITG 120

```
Db 61 QVWFEDQLAOFKMPPTSPECONSCELGIONLSFKVRSQOMLAIIGSSGCGRASLDVITG 120
QY 121 RHGGKINSQGIWINGOPSSPOLYVKCAVAVHQNOLLPNLTVRETLFAIAOMRLPTFS 180
Db 121 RHGGKINSQGIWINGOPSSPOLYVKCAVAVHQNOLLPNLTVRETLFAIAOMRLPTFS 180
QY 181 QOARDKVEDVIAELRLROCADTRVGNMVRGLSGGERRRYSIGVOLLNPGIILDEPT 240
Db 181 QOARDKVEDVIAELRLROCADTRVGNMVRGLSGGERRRYSIGVOLLNPGIILDEPT 240
QY 241 SGLDFTAHNLVKTLSRLAKGNRLVLSIHOPRSDIFRLFDVLVLTSGTPYILCAAOHM 300
Db 241 SGLDFTAHNLVKTLSRLAKGNRLVLSIHOPRSDIFRLFDVLVLTSGTPYILCAAOHM 300
QY 301 VOYFALIGPCPRYSNPADFYVDLTSIDRSREDELTAREKQSLAALFLEKVRDLDEL 360
Db 301 VOYFALIGPCPRYSNPADFYVDLTSIDRSREDELTAREKQSLAALFLEKVRDLDEL 360
QY 361 WKAETKRDLEDTCVSSVTPLDTNCLPSPTKMPGAVQOFTTLIRQISNDFRDLPTLLIH 420
Db 361 WKAETKRDLEDTCVSSVTPLDTNCLPSPTKMPGAVQOFTTLIRQISNDFRDLPTLLIH 420
QY 421 GAELCLMSWTIGELTFHGHSIOLSPMDPAALLFMIGALIPENVILDVISKYSERAMLY 480
Db 421 GAELCLMSWTIGELTFHGHSIOLSPMDPAALLFMIGALIPENVILDVISKYSERAMLY 480
QY 481 ELEDGLYTTGPEFFAKIIGELPEHCAYIIITGMPTMYLANLRPGLOPFLHFLVWLVF 540
Db 481 ELEDGLYTTGPEFFAKIIGELPEHCAYIIITGMPTMYLANLRPGLOPFLHFLVWLVF 540
QY 541 CCRIMALAALALPTFHNASFEFSNALYNSFYLAGFMINLSLMTVPAMISKVSFLRMC 600
Db 541 CCRIMALAALALPTFHNASFEFSNALYNSFYLAGFMINLSLMTVPAMISKVSFLRMC 600
QY 601 ECLMIQFSRRTKPKPLNLTAVSGDKILSMELDSYPLAIVLYIGLSGGEVWLYV 660
Db 601 ECLMIQFSRRTKPKPLNLTAVSGDKILSMELDSYPLAIVLYIGLSGGEVWLYV 660
QY 661 SLRFTKORPSQDM 673
Db 661 SLRFTKORPSQDM 673

RESULT 2
US-10-090-455-7
; Sequence 7, Application US/10090455
; Publication No. US20030027259A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Hongyun
; APPLICANT: Le Bihan, Stephane
; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
; FILE REFERENCE: 100103.406
; CURRENT APPLICATION NUMBER: US/10/090.455
; NUMBER OF SEQ ID NOS: 2002-03-01
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 673
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-090-455-7

Query Match 99.9%; Score 3502; DB 15; Length 673;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 672; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAGRAAERGLPKGATPDGTSGIDRLFFSSSDNSLFTTSGOPNTLEVRDLNTOVDIAS 60
Db 1 MAGRAAERGLPKGATPDGTSGIDRLFFSSSDNSLFTTSGOPNTLEVRDLNTOVDIAS 60
QY 61 QVWFEDQLAOFKMPPTSPECONSCELGIONLSFKVRSQOMLAIIGSSGCGRASLDVITG 120
Db 61 QVWFEDQLAOFKMPPTSPECONSCELGIONLSFKVRSQOMLAIIGSSGCGRASLDVITG 120
```

```
Db 61 QVWFEDQLAOFKMPPTSPECONSCELGIONLSFKVRSQOMLAIIGSSGCGRASLDVITG 120
QY 121 RHGGKINSQGIWINGOPSSPOLYVKCAVAVHQNOLLPNLTVRETLFAIAOMRLPTFS 180
Db 121 RHGGKINSQGIWINGOPSSPOLYVKCAVAVHQNOLLPNLTVRETLFAIAOMRLPTFS 180
QY 181 QOARDKVEDVIAELRLROCADTRVGNMVRGLSGGERRRYSIGVOLLNPGIILDEPT 240
Db 181 QOARDKVEDVIAELRLROCADTRVGNMVRGLSGGERRRYSIGVOLLNPGIILDEPT 240
QY 241 SGLDFTAHNLVKTLSRLAKGNRLVLSIHOPRSDIFRLFDVLVLTSGTPYILCAAOHM 300
Db 241 SGLDFTAHNLVKTLSRLAKGNRLVLSIHOPRSDIFRLFDVLVLTSGTPYILCAAOHM 300
QY 301 VOYFALIGPCPRYSNPADFYVDLTSIDRSREDELTAREKQSLAALFLEKVRDLDEL 360
Db 301 VOYFALIGPCPRYSNPADFYVDLTSIDRSREDELTAREKQSLAALFLEKVRDLDEL 360
QY 361 WKAETKRDLEDTCVSSVTPLDTNCLPSPTKMPGAVQOFTTLIRQISNDFRDLPTLLIH 420
Db 361 WKAETKRDLEDTCVSSVTPLDTNCLPSPTKMPGAVQOFTTLIRQISNDFRDLPTLLIH 420
QY 421 GAELCLMSWTIGELTFHGHSIOLSPMDPAALLFMIGALIPENVILDVISKYSERAMLY 480
Db 421 GAELCLMSWTIGELTFHGHSIOLSPMDPAALLFMIGALIPENVILDVISKYSERAMLY 480
QY 481 ELEDGLYTTGPEFFAKIIGELPEHCAYIIITGMPTMYLANLRPGLOPFLHFLVWLVF 540
Db 481 ELEDGLYTTGPEFFAKIIGELPEHCAYIIITGMPTMYLANLRPGLOPFLHFLVWLVF 540
QY 541 CCRIMALAALALPTFHNASFEFSNALYNSFYLAGFMINLSLMTVPAMISKVSFLRMC 600
Db 541 CCRIMALAALALPTFHNASFEFSNALYNSFYLAGFMINLSLMTVPAMISKVSFLRMC 600
QY 601 ECLMIQFSRRTKPKPLNLTAVSGDKILSMELDSYPLAIVLYIGLSGGEVWLYV 660
Db 601 ECLMIQFSRRTKPKPLNLTAVSGDKILSMELDSYPLAIVLYIGLSGGEVWLYV 660
QY 661 SLRFTKORPSQDM 673
Db 661 SLRFTKORPSQDM 673
```

```
RESULT 3
US-09-989-981A-4
; Sequence 4, Application US/0998981A
; Publication No. US20030049730A1
; GENERAL INFORMATION:
; APPLICANT: Hobbs, Helen H.
; APPLICANT: Shan, Bel
; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989.981A
; PRIOR APPLICATION NUMBER: US 60/252.235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253.645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 672
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: mouse ABCG8 (mABCG8)
US-09-989-981A-4

Query Match 82.2%; Score 2883.5; DB 11; Length 672;
```

Best Local Similarity 81.9%; Pred. No. 6.5e-283;
Matches 551; Conservative 52; Mismatches 69; Indels 1; Gaps 1;

```

QY 1 MAGRAAERGLPKGATPQDTSGLDRLFSESDNSLFTYSGOPNTLEVRDLNVOVLAS 60
D 1 MAEKTEETOLMNTVLDASGLDLSFSESDNSLFTYSGOSNTLEVRDLNVOVLAS 60
QY 61 QVWFEOLOAFKIPMRSHSSQDSELCIRMLSEKVRSGOMLAIIGSSGCRASLIDVTG 120
D 61 QVWFEOLOAFKIPMRSHSSQDSELCIRMLSEKVRSGOMLAIIGSSGCRASLIDVTG 120
QY 121 RGHGKIKSGOIMYNGPSSPOLYRKCAVAHROHNLPLVRETAFIAQMLPTFS 180
D 121 RGHGKIKSGOIMYNGPSSPOLYRKCAVAHROHNLPLVRETAFIAQMLPTFS 180
QY 181 QAOARKREVEDVIAELRLROCADTRVGMVYRGSGGERRRVSIGVOLLNMPGILLIDEP 240
D 181 QAOARKREVEDVIAELRLROCADTRVGMVYRGSGGERRRVSIGVOLLNMPGILLIDEP 240
QY 241 SGIDSTFANHLVTLRLAKGNRLVLSLHQPBSDIFRLFDVLVLTSGTPIYLGAQOM 300
D 241 SGIDSTFANHLVTLRLAKGNRLVLSLHQPBSDIFRLFDVLVLTSGTPIYLGAQOM 300
QY 301 VOYFTAGVPCRYSPNADRYVDTLSIDRSRQDELATREKASLALFEKYRDLDDEL 360
D 301 VOYFTAGVPCRYSPNADRYVDTLSIDRSRQDELATREKASLALFEKYRDLDDEL 360
QY 361 WKAETDLEDTCVSSVTPIDTNCLEPPTKMGAVOQFTLLRLROISNDFRDLPTLTH 420
D 361 WKAETDLEDTCVSSVTPIDTNCLEPPTKMGAVOQFTLLRLROISNDFRDLPTLTH 420
QY 421 GAACLSMTIGLYGHGSIQSLSEMDTALLFMIGALIPFNVLIDVISCYERAMLY 480
D 421 GAACLSMTIGLYGHGSIQSLSEMDTALLFMIGALIPFNVLIDVISCYERAMLY 480
QY 481 ELEDGTYTGVPYFAATIGELPEHCAYIIYIGPPTVLANLRGLOPFLHLLVLYV 540
D 481 ELEDGTYTGVPYFAATIGELPEHCAYIIYIGPPTVLANLRGLOPFLHLLVLYV 540
QY 541 CCRIMAAALPFRHMASFSPNALYNSFYLAGGPMNLSSIMTPANISRYSELRMCE 600
D 541 CCRIMAAALPFRHMASFSPNALYNSFYLAGGPMNLSSIMTPANISRYSELRMCE 600
QY 601 BELAMIOESRRTYKPLGNLTAVSGDKILSAMELDYPYLAIIYLIVIGLSGFVLYYL 660
D 601 BELAMIOESRRTYKPLGNLTAVSGDKILSAMELDYPYLAIIYLIVIGLSGFVLYYL 660
QY 661 SRFKOKRSQDM 673
D 661 SRFKOKRSQDM 673
QY 660 SGLMOIOFNGHLYTQIGNFTESIIGDTWISAMDLNHLVLAIIYLIVIGLSGFVLYYL 659
D 660 SGLMOIOFNGHLYTQIGNFTESIIGDTWISAMDLNHLVLAIIYLIVIGLSGFVLYYL 659

```

RESULT 4
US-09-837-992-3

Sequence 3, Application US/09837992
Patent No. US2002008167A1

GENERAL INFORMATION:
APPLICANT: Tian, Hui

APPLICANT: Schultz, Joshua

APPLICANT: Shan, Bei

APPLICANT: Tularik Inc.

TITLE OF INVENTION: Sltosterolemia Susceptibility Gene (SSG): Compositions

FILE REFERENCE: 018781-00602005

CURRENT APPLICATION NUMBER: US/09/837, 992

PRIOR APPLICATION NUMBER: US 60/198, 465

PRIOR FILING DATE: 2000-04-18

PRIOR APPLICATION NUMBER: US 60/204, 234

PRIOR FILING DATE: 2000-05-15

NUMBER OF SEQ ID NOS: 45

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 3

LENGTH: 651
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: human sltosterolemia susceptibility gene (SSG)
OTHER INFORMATION: amino acid sequence
US-09-837-992-3

Query Match 19.9%; Score 697; DB 9; Length 651;
Best Local Similarity 28.9%; Pred. No. 2.4e-61;
Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;

```

QY 16 TPQDTSGIDRLFSESDNSLFTYSGOPNTLEVRDLNVOVLASQVWFEOLOAFKIPMR 75
D 16 TPQDTSGIDRLFSESDNSLFTYSGOPNTLEVRDLNVOVLASQVWFEOLOAFKIPMR 75
QY 76 TSPSCONSCIELGIONLSFKYRSGOMLAIIGSSGCRASLIDVTGR-GHGGKIKSGOIWI 134
D 76 TSPSCONSCIELGIONLSFKYRSGOMLAIIGSSGCRASLIDVTGR-GHGGKIKSGOIWI 134
QY 135 NGQPSPOLYRKCAVAHROHNLPLVRETAFIAQMLPTFSQAOARKREVEDVIAE 194
D 135 NGQPSPOLYRKCAVAHROHNLPLVRETAFIAQMLPTFSQAOARKREVEDVIAE 194
QY 195 LRLROCADTRVGMVYRGSGGERRRVSIGVOLLNMPGILLIDEPSTGIDSTFANHLVLT 254
D 195 LRLROCADTRVGMVYRGSGGERRRVSIGVOLLNMPGILLIDEPSTGIDSTFANHLVLT 254
QY 255 LSRIKAGNRLVLSLHQPBSDIFRLFDVLVLTSGTPIYLGAQOMVOYFTAGVPCRY 314
D 255 LSRIKAGNRLVLSLHQPBSDIFRLFDVLVLTSGTPIYLGAQOMVOYFTAGVPCRY 314
QY 315 SNADRYVDTLSIDRSRQDELATREKASLALFEKYRDLDDEL 362
D 315 SNADRYVDTLSIDRSRQDELATREKASLALFEKYRDLDDEL 362
QY 363 AETKRDLEDTCVSSVTPIDTNCLEPPTKMGAVOQFTLLRLROISNDFRDLPTLTH 421
D 363 AETKRDLEDTCVSSVTPIDTNCLEPPTKMGAVOQFTLLRLROISNDFRDLPTLTH 421
QY 422 AENCLSMITIGLYFPG-----HGSITLSEMDTALLFMIGALIPFNVLIDVISCYER 475
D 422 AENCLSMITIGLYFPG-----HGSITLSEMDTALLFMIGALIPFNVLIDVISCYER 475
QY 476 AMLYLEGLTYTGVEFAKILGELPEHCAYIIYIGPPTVLANLRGLOPFLHLLVLYL 535
D 476 AMLYLEGLTYTGVEFAKILGELPEHCAYIIYIGPPTVLANLRGLOPFLHLLVLYL 535
QY 536 WLVPFCRIMAAALPFRHMASFSPNALYNSFYLAGGPMNLSSIMTPANISRYSELRMCE 577
D 536 WLVPFCRIMAAALPFRHMASFSPNALYNSFYLAGGPMNLSSIMTPANISRYSELRMCE 577
QY 578 INLSIMTPANISRYSELRMCEGLMCKIOFSRRTYKPLGNLTAVS 625
D 578 INLSIMTPANISRYSELRMCEGLMCKIOFSRRTYKPLGNLTAVS 625
QY 550 RNIQEMPIPKIISYFTFOKYCSLIVNEFGLNFTCGSSNVSATTN 597
D 550 RNIQEMPIPKIISYFTFOKYCSLIVNEFGLNFTCGSSNVSATTN 597

```

RESULT 5
US-09-989-981A-6

Sequence 6, Application US/09989981A
Publication No. US20030049730A1

GENERAL INFORMATION:
APPLICANT: Hobbs, Helen H.

APPLICANT: Shan, Bei

APPLICANT: Barnes, Robert

APPLICANT: Tularik Inc.

TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use

FILE REFERENCE: 018781-00732005

CURRENT APPLICATION NUMBER: US/09/989, 981A

PRIOR FILING DATE: 2002-07-23

PRIOR APPLICATION NUMBER: US 60/252, 235

RESULT 6
US-10-090-455-6
Sequence 6, Application US/10090455
Publication No. US20030027259A1
GENERAL INFORMATION:
APPLICANT: Chen, Hongyun
APPLICANT: Le Bihan, Stephane
TITLE OF INVENTION: NOVEL ABCq TRANSPORTER AND USES THEREOF
FILE REFERENCE: 100103.406
CURRENT APPLICATION NUMBER: US/10/090,455
CURRENT FILING DATE: 2002-03-01

```

QY      16  TPDDTSGLDRLEFSSSDNSLFTYSGQONTLEVDNLVQYDLAQVMPQLQFKPM  75
Db      8  TPQSGMCLQVNRSSOSLEGAPAT-APRPHSIGILHASTYSHRRK-PWMD-TTSCROOM  64
QY      76  TSPSCONSELGIQNLSTFVRSGQMALAIIGSSGCRASLLVYTGR-GHGKIRKGOIWI  134
Db      65  TROI-----LKDVLSYVESGQIMCIGSSGSGKTYLLDMSGRIGRAFT-IGEYV  118
QY      135  NGOPSSPOLYRKCVANRHONHNLNLYRELIATIAOMRPLRPFSSQNRKREDVIAE  199
Db      116  NGRALRREPQDPQFSTVYLDSDTLSSLYRELIHTALLAT-RGNPESFORKEAVMAE  174
QY      195  LRLRUCADTRVGNMYRNGVSGGERRRVSGIYOLAMPILILDEPTSGSTFHAHLVKT  255
Db      175  LSTSHVADRILGNYSIGISGTGBERRRVSIAQLDDPVMFLDEPTTGLDGMTANQIVYL  234
QY      255  LSLTAGNRHLYLISLQRPDSDFRFLDVLMTSGTPYILCAAOHMOYFATGICPCPRY  314
Db      235  LVELARRNRKRYVLTTHQRPSELFQLEFDKITALISGELIFCGTPAEMDFENDCGPCPEH  294
QY      315  SNPADEYVDLTSIDRSRQEOLATREKASLALP-----LEKRDQDDPLMK  362
Db      295  SNPFDEYMDLTVDQSKREIETESKRYQMIESAKKAKICHKTLKNIERKHL-----  348
QY      363  AETKQDEDETCVESSSTPLDTNCLSPRK-MPQAVQFTTLIRROISNDFDLPTLIHG  421
Db      349  -----KTLPM-----VFKRKDSGVFSKIGVLLRRYTRNLVNNKLAIVTRL  390
QY      422  AEACLSMTLIGLYEFG-----HSGIQLSFMDTALLPMICALPFVNIIDVISCYSER  475
Db      391  LQHLINGLFLFLFVLRVNSVLKAGIQ-----DRGILLYQFGALRPYTGMLNANVLFYLR  446
QY      476  AMLYELEDGLYTTGCFYPAKILGELPEHCAYIIHYGPTYWLANLRPGLOPFLHFLV  535
Db      447  AVSDOSQGLQYOKOMQMLAYVALHVLPEFSVATMIFSSVCYWTGILPEVARF-----  489
QY      536  WLVPFCRRIMATAAALALPTFHMASEFS-----NALYNSPYLAG-----GFM  577
Db      500  -----GFSMAILLAPHLIGELFLVYLGLIYQNPNIYNSVALISTINGVLVSGFL  549
QY      578  INLSLMTVPAMISKVSFLRMCFEGLMKIOPSRRTKKPLDGLNTIAVS  625
Db      550  RNIOEMPRIPEKIIISYTFQKRYCSEILVYVNEFGLNFTGGSNVSATVIN  597

```

RESULT 7
US-09-837-992-1
Sequence 1, Application US/09837992
Patent No. US20020081687A1
GENERAL INFORMATION:
APPLICANT: Tian, Hui
APPLICANT: Schultz, Joshua
APPLICANT: Shan, Bei
APPLICANT: Tularik Inc.
TITLE OF INVENTION: Sticosterolemia Susceptibility Gene (SSG): Compositions
TITLE OF INVENTION: and Methods of Use
FILE REFERENCE: 018781-0060200S
CURRENT APPLICATION NUMBER: US/09/837,992
CURRENT FILING DATE: 2001-04-18
PRIORITY APPLICATION NUMBER: US 60/198,465
PRIORITY FILING DATE: 2000-04-18

```

; PRIOR APPLICATION NUMBER: US 60/204,234
; PRIOR FILING DATE: 2000-05-15
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 1
; LENGTH: 652
; TYPE: PR
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: mouse sitosterolemia susceptibility gene (SSG)
; OTHER INFORMATION: amino acid sequence
US-09-837-992-1

```

```

Query Match      19.6%; Score 688.5; DB 9; Length 652;
Best Local Similarity 28.1%; Pred. No. 1.7e-60;
Matches 188; Conservative 125; Mismatches 233; Indels 123; Gaps 16;

```

```

OY 45 NTLEVRDLNVOYDLASQV-PWFEQLAQFMPTSPSCNSCEIGI-QNLSFKVRSQOMLA 102
DB 37 HSLGVLHVSYV-SNRVGPW-----WNKSCQCKMROLDKVDLSVIESQIMC 84
OY 103 IIGSSGGRASLLDVTGGHGGKIKSGQIMWINGQSSQVLYRKCAVHRQNNQLPNT 162
DB 85 IIGSSGSKTTLDDAISGRRLRTGLGGEVFNVCGLRDQFQDFSYVLQSDVFLSLT 144
OY 163 VRETLAFLAQMRLPRTFSQADRKVEDYIAELRLQCADTRVNMNYVGLSGGERRRS 222
DB 145 VRETLAFLAQMRLPRTFSQADRKVEDYIAELRLQCADTRVNMNYVGLSGGERRRS 203
OY 223 IGVOLLNMGILLDEPTSGDLSFTAHNLVKTLSRLAKGNRLVLSLHQPBSDIRFLDL 282
DB 204 IAAQLDQPKVMMLDEPTGGLDQMTANQVLLLAELARDRIYVITIQHPSRELQHQHDK 263
OY 283 VLLMTSGPPIYIGAQOHMVOYTAIGYPCPRPSNADYVDLTSIDRSREDELTATREKA 342
DB 264 IAILTYGELVFCGTEPEMLGFEFNCGYPCPEHSNPFDEYMDLTSVDTOSREREIETRYKV 323
OY 343 OSIALF-----LEKVRDLDFLMAETKRDDEPTCVSSVTPPLDTNCLPSP 390
DB 324 QMLECAFKESDIYHKILENIERAYL-----KILPM---VPEKT 359
OY 391 K-MGAVOQFTLLIRROISNDFRDLPTLLIHGAECALMSMTIGF--LYFGHSIOLSPND 447
DB 360 KDPGMFGKIGVLLRRVTRNLRNRKQAVIMRLVQMLINGFLIFLILRVQNNLTGKGAQD 419
OY 448 TALLFMIGALIPFENVIIDVYSKCSERAMLYELEDGLYTGPFYFAKILGELPERCAY 507
DB 420 RVGLLYQVLGATPYGMNAVANLFPMLRAVSDQESQDGLYHKQMILLAVLVLPFSVIA 479
OY 508 ILYGMPTMYMLANLRPGLOPFLHFLVWLVLVFCRIMALAAALLPFIHMASPFSNML- 566
DB 480 TVIFSSVCYWTGLGTPVARF-----GYSAALLAHLIGEEFLTLVL 522
OY 567 -----YNSFLAGFMINLSLMTVPAMISKVSFLRCEFGMLKIOFS 609
DB 523 GIVQNPNIIVNSIVALLSISGLLIGSGFIRNIOEMPIPKILGYTFPQYCCCEILVNF- 581
OY 610 RRTYKMPIGNLTIVASGDKIISAMELDSYPLXAI-----YLIYI 648
DB 582 ---YGL---NFTCGSNTSM-----NHPMCALITQGVQFIEKTCPGATSRTANFLILY 629
OY 649 GLSGGFVYL 657
DB 630 GFIPALVYL 638

```

```

RESULT 8
US-09-989-981A-2
; Sequence 2, Application US/09989981A
; Publication No. US20030049730A1
; GENERAL INFORMATION:
; APPLICANT: Hobbs, Helen H.
; APPLICANT: Shan, Bei

```

```

; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 2
; LENGTH: 652
; TYPE: PR
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: mouse ABCG5 (mABCG5)
US-09-989-981A-2

```

```

Query Match      19.6%; Score 688.5; DB 11; Length 652;
Best Local Similarity 28.1%; Pred. No. 1.7e-60;
Matches 188; Conservative 125; Mismatches 233; Indels 123; Gaps 16;

```

```

OY 45 NTLEVRDLNVOYDLASQV-PWFEQLAQFMPTSPSCNSCEIGI-QNLSFKVRSQOMLA 102
DB 37 HSLGVLHVSYV-SNRVGPW-----WNKSCQCKMROLDKVDLSVIESQIMC 84
OY 103 IIGSSGGRASLLDVTGGHGGKIKSGQIMWINGQSSQVLYRKCAVHRQNNQLPNT 162
DB 85 IIGSSGSKTTLDDAISGRRLRTGLGGEVFNVCGLRDQFQDFSYVLQSDVFLSLT 144
OY 163 VRETLAFLAQMRLPRTFSQADRKVEDYIAELRLQCADTRVNMNYVGLSGGERRRS 222
DB 145 VRETLAFLAQMRLPRTFSQADRKVEDYIAELRLQCADTRVNMNYVGLSGGERRRS 203
OY 223 IGVOLLNMGILLDEPTSGDLSFTAHNLVKTLSRLAKGNRLVLSLHQPBSDIRFLDL 282
DB 204 IAAQLDQPKVMMLDEPTGGLDQMTANQVLLLAELARDRIYVITIQHPSRELQHQHDK 263
OY 283 VLLMTSGPPIYIGAQOHMVOYTAIGYPCPRPSNADYVDLTSIDRSREDELTATREKA 342
DB 264 IAILTYGELVFCGTEPEMLGFEFNCGYPCPEHSNPFDEYMDLTSVDTOSREREIETRYKV 323
OY 343 OSIALF-----LEKVRDLDFLMAETKRDDEPTCVSSVTPPLDTNCLPSP 390
DB 324 QMLECAFKESDIYHKILENIERAYL-----KILPM---VPEKT 359
OY 391 K-MGAVOQFTLLIRROISNDFRDLPTLLIHGAECALMSMTIGF--LYFGHSIOLSPND 447
DB 360 KDPGMFGKIGVLLRRVTRNLRNRKQAVIMRLVQMLINGFLIFLILRVQNNLTGKGAQD 419
OY 448 TALLFMIGALIPFENVIIDVYSKCSERAMLYELEDGLYTGPFYFAKILGELPERCAY 507
DB 420 RVGLLYQVLGATPYGMNAVANLFPMLRAVSDQESQDGLYHKQMILLAVLVLPFSVIA 479
OY 508 ILYGMPTMYMLANLRPGLOPFLHFLVWLVLVFCRIMALAAALLPFIHMASPFSNML- 566
DB 480 TVIFSSVCYWTGLGTPVARF-----GYSAALLAHLIGEEFLTLVL 522
OY 567 -----YNSFLAGFMINLSLMTVPAMISKVSFLRCEFGMLKIOFS 609
DB 523 GIVQNPNIIVNSIVALLSISGLLIGSGFIRNIOEMPIPKILGYTFPQYCCCEILVNF- 581
OY 610 RRTYKMPIGNLTIVASGDKIISAMELDSYPLXAI-----YLIYI 648
DB 582 ---YGL---NFTCGSNTSM-----NHPMCALITQGVQFIEKTCPGATSRTANFLILY 629
OY 649 GLSGGFVYL 657
DB 630 GFIPALVYL 638

```


Db 507 FMMETLMMVAVSASSMALAIAAGOSVSVATLMTICVFMMIFSGLLVNLTTIASWLS 566
Oy 589 WISKVSFLRMCEGLMKIQFSRRTYKMPIGNLT-----IAVSGDKIL--SAMELDSYP 639
Db 567 WLOYSIPRYGFTALQHNHEFLGQNF--CPGLNATGNPCNYATCGEELVYAGIDLSRWG 625
Oy 640 LYAILIYIGLSGGMVLYVYSLRPIKQ 667
Db 626 LMKNHVALACMIVIFLTAYIKLLEFLK 653

RESULT 13

US-09-961-086-1
; Sequence 1, Application US/09961086
; Publication No. US20030036645A1
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF MARYLAND, BALTIMORE
; APPLICANT: ROSS, Douglas D.
; APPLICANT: DOYLE, L. Austin
; APPLICANT: ABRUZZO, Lynne
; TITLE OF INVENTION: BREAST CANCER RESISTANCE PROTEIN (BCRP) AND THE DNA
; TITLE OF INVENTION: WHICH ENCODES IT
; FILE REFERENCE: EPI9376-019
; CURRENT APPLICATION NUMBER: US/09/961,086
; PRIOR FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/073,763
; PRIOR FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: PCT/US99/02577
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-961-086-1

Query Match 18.3%; Score 640.5; DB 11; Length 655;
Best Local Similarity 27.2%; Pred. No. 1.3e-55;
Matches 187; Conservative 139; Mismatches 273; Indels 89; Gaps 21;

Oy 19 DTSGIQDLRFSSSEDSNLSLYFTYSGOPNTLEVRDNLVYDLSQVWPFQIAQFKMPWTSP 78
Db 16 NTNG-----FPATASNDLKATTEGA--VLSFHNITRYKRLKSGF-----LP----- 54
Oy 79 SCQNSCELG I-QNLSFKVRSQGMALIISSGCGRASLDVYTRGHHGKIKSGQIWMNGQ 137
Db 55 -CRKPEKEILSNINGIMKPG-LNALILGPTGGKSSLDVLAARDPGSL--SGDVLINGA 111
Oy 138 PSSQOLVKKC-VAHYRQHNQLLPNLTVRETLAFIAOMRLPRTFSQAOQRKVEDVIAELR 196
Db 112 PRPANE--KCNNGYVODVWVGTLTRENLQFSALRLATMTMHEKNERINRYIQELG 169
Oy 197 LKOCADTVNGMYVNGSGGERRRYSIGVQLMNGIILDEPTSGDLSFAHNLVKTLS 256
Db 170 LDKVADSVMGTQIRGVSGERKRKRSIGMELTDPSSILFDEPTTGSDLSANAVLLLK 229
Oy 257 RLAKGNRLVLSLHOPRSDFRLPDLVLLMTSGPTIYGAOAHMVOYTAIGYPCPRYSN 316
Db 230 RMSKGRITLIFSHPRIKFLPDLSTLLASGRLMFGHPQOEAALGYESAGYHCEAVNN 289
Oy 317 PADFVVDLTSIDR-----SREQLATRE--KAQSLAALFLEKVRDL--DDELMAETK-- 366
Db 290 PADFVVDLTSIDR-----SREQLATRE--KAQSLAALFLEKVRDL--DDELMAETK-- 366
Oy 367 -----DLDEDTYCVSSVTPDLTNCPLSPKMPGAVOQFTLLIRQIJSNDRDLPTLLI 419
Db 349 LHLQSGGKKKKITVFEKISYTTSPC-----HQLRWVSKRSFKNLGNPOASIA 397
Oy 420 HGAERCLMSMTIGFLYFGHSIQLSFMDTALLFMIGALLFPNVLIDYISCYG----- 473
Db 398 QIIVTVGLVIGALYIGLKNDSYTGIONRAGVLEFL-----TTNCFSSVSVAVE 446

Oy 474 ----BRALVYELDEGLTTCGPYFAKILGE-LPEHCAYIIYGMPTVLANLRGLOPF 528
Db 447 LFVVEKRLFIHEYISGYRVSSYFLGKILSDLPMTMPSIFTCIYVFMGLKRAQAF 506
Oy 529 ILHFLVLMVYFCCRIAMALAAALPTMHASFFESNALYNSPYLAGGMINLSLWTPA 588
Db 507 FMMETLMMVAVSASSMALAIAAGOSVSVATLMTICVFMMIFSGLLVNLTTIASWLS 566
Oy 589 WISKVSFLRMCEGLMKIQFSRRTYKMPIGNLT-----IAVSGDKIL--SAMELDSYP 639
Db 567 WLOYSIPRYGFTALQHNHEFLGQNF--CPGLNATGNPCNYATCGEELVYAGIDLSRWG 625
Oy 640 LYAILIYIGLSGGMVLYVYSLRPIKQ 667
Db 626 LMKNHVALACMIVIFLTAYIKLLEFLK 653

RESULT 14

US-09-866-866A-10
; Sequence 10, Application US/09866866A
; Patent No. US2002010224A1
; GENERAL INFORMATION:
; APPLICANT: Sorrentino, Brian
; APPLICANT: Schuetz, John
; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
; FILE REFERENCE: 1340-1-021CIP2
; CURRENT APPLICATION NUMBER: US/09/866,866A
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: 09/584,586
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: PCT/US99/11825
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: 60/086,988
; PRIOR FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-866-866A-10

Query Match 18.2%; Score 638.5; DB 10; Length 655;
Best Local Similarity 27.9%; Pred. No. 2e-55;
Matches 175; Conservative 131; Mismatches 254; Indels 67; Gaps 17;

Oy 80 CONCELG I-QNLSFKVRSQGMALIISSGCGRASLDVYTRGHHGKIKSGQIWMNGQ 138
Db 55 -CRKPEKEILSNINGIMKPG-LNALILGPTGGKSSLDVLAARDPGSL--SGDVLINGA 112
Oy 139 SSPOLVKKC-VAHYRQHNQLLPNLTVRETLAFIAOMRLPRTFSQAOQRKVEDVIAELR 197
Db 113 PRPANE--KCNNGYVODVWVGTLTRENLQFSALRLATMTMHEKNERINRYIEELG 170
Oy 198 RQCADTVNGMYVNGSGGERRRYSIGVQLMNGIILDEPTSGDLSFAHNLVKTLS 257
Db 171 DKVADSKVGTQIRGVSGERKRKRSIGMELTDPSSILFDEPTTGSDLSANAVLLLK 230
Oy 258 LAKGNRLVLSLHOPRSDFRLPDLVLLMTSGPTIYGAOAHMVOYTAIGYPCPRYSN 317
Db 231 MSKGRITLIFSHPRIKFLPDLSTLLASGRLMFGHPQOEAALGYESAGYHCEAVNN 289
Oy 318 ADPFDLTSIDR-----SREQLATRE--KAQSLAALFLEKVRDL--DDELMAETK-- 366
Db 291 ADPFDLTSIDR-----SREQLATRE--KAQSLAALFLEKVRDL--DDELMAETK-- 366
Oy 367 -----DLDEDTYCVSSVTPDLTNCPLSPKMPGAVOQFTLLIRQIJSNDRDLPTLLI 420
Db 350 HQLSGGKKKKITVFEKISYTTSPC-----HQLRWVSKRSFKNLGNPOASIAQ 398
Oy 421 GAEACLSMTIGFLYFGHSIQLSFMDTALLFMIGALLFPNVLIDYISCYG----- 473

```

Db      399 IIVTVGLVIGAIYFGLKNDSTGIONRAGVLEFL-----TTNOCSSVSAGVEL 447
QY      474 ---ERAMLYELEDGLYTTGPYEFKILCE-LPEHCAYIIITGCMPTWLANLRPGLOPEL 529
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      448 FVVEKELFHEIYISGYRVSSTYFLGKLDLPMRLPSTIFCIYFMGLKPKADAEF 507
QY      530 LHEFLVWLVECCRIALAAALLPTFHNASFEFSNALYNSFYLAGGFMINLSLMTVPAM 589
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      508 VMHFTLMVAVSASSMALAIAAGQSVSVATILMTICFVPMHIFSGLLVNLTTIASWLSM 567
QY      590 ISKVSFLRWCCEBIMKIOFSRRYKRPGLNLT-----IANSCKIL--SAMELDSYPL 640
Db      568 LQFSIPRYGFTALQHNELGONF-CPGLNATGNMPCNYATCTGEEYLVKQGIDLSPMGL 626
QY      641 VAIYLVIGLSGGFMVLYVSLRFIRK 667
Db      627 WKNHVALACMIYIFLTIAVTKLFLKK 653

```

```

QY      530 LHEFLVWLVECCRIALAAALLPTFHNASFEFSNALYNSFYLAGGFMINLSLMTVPAM 589
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      508 VMHFTLMVAVSASSMALAIAAGQSVSVATILMTICFVPMHIFSGLLVNLTTIASWLSM 567
QY      590 ISKVSFLRWCCEBIMKIOFSRRYKRPGLNLT-----IANSCKIL--SAMELDSYPL 640
Db      568 LQFSIPRYGFTALQHNELGONF-CPGLNATGNMPCNYATCTGEEYLVKQGIDLSPMGL 626
QY      641 VAIYLVIGLSGGFMVLYVSLRFIRK 667
Db      627 WKNHVALACMIYIFLTIAVTKLFLKK 653

```

Search completed: July 25, 2003, 17:23:59
Job time : 57 secs

```

RESULT 15
US-10-090-455-5
: Sequence 5, Application US/10090455
: Publication No. US20030027259A1
: GENERAL INFORMATION:
: APPLICANT: Chen, Hongyun
: APPLICANT: Le Bihan, Stephane
: TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
: FILE REFERENCE: 100103.406
: CURRENT APPLICATION NUMBER: US/10/090,455
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 5
: LENGTH: 655
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-090-455-5

```

Query Match 18.2%; Score 638.5; DB 15; Length 655;
Best Local Similarity 27.9%; Pred. No. 2e-55;
Matches 175; Conservative 131; Mismatches 254; Indels 67; Gaps 17;

```

QY      80 CONSCELGI-QNLSFVRSQGMALIGSSGCRASILDVITGRGHGKIKSGQIWINOP 138
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      55 CRKPYEKELSLININGIMPGL-LMAILGPTGGKSSLDVLAARKDPSGL-SGDVLTINGAP 112
QY      139 SSPOLVRKC-VAHVROHNOQLPNTLVRETLAFIAQMLPRTFSQAORDKREVDYIAELRL 197
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      113 RPANF-KCNSGYVQDVVGTIVRENLQFSALRLATWTNHEKNERINRVIEELGL 170
QY      198 ROCADTRYGNMYVRSLSGERRRVSIGQLMNPGLILDEPTSGDSTFAHNLVKTISR 257
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      171 DVADSKVGTQPIKVSIGERRKRTSIMGELTYDPSILSDEPTGIDSSSTANAAVLLILKR 230
QY      258 LAKGNRLVLISLHOPRSDIFRLFDVLMTSGTPYILGAQHMOGYFALIGYPCPRYSNP 317
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      231 MSKQRTIIFSIHQPRYSIFKLFDSLTILASGRIMFHPAOEALGYEFESAGYHCEAYNPP 290
QY      318 ADYVYDLSIDRR---SREOLATRE--KAOSLAALFLEKVRDL--DDFLMKAETK--- 366
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      291 ADFEIDINDSTAVALNRDEDFKATEIIEPSKODKPLIEKLAELIYVNSFYK-ETKREL 349
QY      367 -----DLDEPTCYESSYTPLDITNCLESPYTMPCGAVOQFTLLIRQISNDRDLPILLIH 420
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      350 HOLSGGEKKRITVFEKELSYTSTFC-----HQLRWYSKRSEKNILGNPOASIAQ 398
QY      421 GAELCLMSMTIGFLYFGHSIOLSPMDTALLFMIGALIPFNVLIDVSKYS----- 473
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      399 IIVTVGLVIGAIYFGLKNDSTGIONRAGVLEFL-----TTNOCSSVSAGVEL 447
QY      474 ---ERAMLYELEDGLYTTGPYEFKILCE-LPEHCAYIIITGCMPTWLANLRPGLOPEL 529
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      448 FVVEKELFHEIYISGYRVSSTYFLGKLDLPMRLPSTIFCIYFMGLKPKADAEF 507

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 17:07:19 ; Search time 26 Seconds
(without alignments)
1217.268 Million cell updates/sec

Title: US-09-989-981a-8
Perfect score: 3506
Sequence: 1 MACKAAEERGLPKGATPDDT.....FVLYVYSLKFIKPSQDW 673

Scoring table: BLAST62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3502	99.9	673	1	ABG8_HUMAN
2	2873	81.9	673	1	ABG8_MOUSE
3	2829.5	80.7	672	1	ABG8_RAT
4	713	20.3	652	1	ABG5_RAT
5	697	19.9	651	1	ABG5_HUMAN
6	691.5	19.7	652	1	ABG5_MOUSE
7	656	18.7	687	1	WHIT_DROME
8	653	18.6	1294	1	YOH5_YEAST
9	640.5	18.3	695	1	ABG2_HUMAN
10	627	17.9	695	1	WHIT_ANGGA
11	623.5	17.8	679	1	WHIT_CERCA
12	621	17.7	666	1	ABG1_MOUSE
13	620.5	17.7	677	1	WHIT_LUCCU
14	617	17.6	678	1	ABG1_HUMAN
15	600	17.1	598	1	YPC3_CAEEL
16	583	16.6	709	1	WHIT_ANGAL
17	573.5	16.4	646	1	ABG4_HUMAN
18	562.5	16.0	1049	1	ADP1_YEAST
19	552	15.7	666	1	SCRT_DROME
20	511	14.6	610	1	YOS3_CAEEL
21	464	13.2	1564	1	BROW_YEAST
22	463.5	13.2	675	1	BROW_DROME
23	452.5	12.9	650	1	ABG3_MOUSE
24	437	12.5	668	1	BROW_PROVI
25	434.5	12.4	1499	1	CDR2_CANAL
26	431	12.3	1529	1	PDRT_YEAST
27	424.5	12.1	1501	1	CDR1_CANAL
28	412	11.8	1490	1	CDR4_CANAL
29	401	11.4	1333	1	YNG9_YEAST
30	397	11.3	1501	1	SNQ2_YEAST
31	388.5	11.1	1501	1	CDR3_CANAL
32	388.5	11.1	1501	1	BFR1_SCHPO
33	385	11.0	1511	1	PDR5_YEAST

RESULT 1	ABG8_HUMAN	STANDARD:	PRT:	673 AA.	
34	349.5	10.0	1511	1	PDR5_YEAST
35	333	9.5	1410	1	PDR5_YEAST
36	270.5	7.7	670	1	NRTC_SYNY3
37	252.5	7.2	894	1	YH1H_ECOLI
38	251	7.2	371	1	Y4OS_RHISN
39	250	7.1	1704	1	ABG3_HUMAN
40	248.5	7.1	355	1	CYSA_SYNY3
41	248	7.1	352	1	AGLK_RHIME
42	244	7.0	659	1	NRTC_SYNP7
43	241.5	6.9	332	1	SMOK_RHOSH
44	241	6.9	236	1	CYSA_CHLYU
45	238.5	6.8	344	1	CYSA_SYNP7

ALIGNMENTS:

ABG8_HUMAN	STANDARD:	PRT:	673 AA.
OC	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DE	15-SEP-2003 (Rel. 42, Last annotation update)		
GN	ATP-binding cassette, sub-family G, member 8 (Sterolin-2).		
OS	ABG8.		
OC	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
ON	[1]		
RP	SEQUENCE FROM N.A., VARIANTS STIGSTEROLEMIA THR-231; GLN-263; ARG-574		
RP	AND ARG-596; AND VARIANT CYS-54;		
RX	MEDLINE-20553648; PubMed-11099417;		
RA	Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,		
RA	Kwiterovich P., Shan B., Barnes R., Hobbs H.H.,		
RT	"Accumulation of dietary cholesterol in sitosterolemia caused by		
RT	mutations in adjacent ABC transporters.";		
RL	Science 290:1771-1775(2000).		
RN	[2]		
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), VARIANTS STIGSTEROLEMIA		
RP	HIS-184; THR-231; GLN-263; HIS-405; PRO-501; SER-543; PRO-572;		
RP	GLU-574; ARG-574; ARG-596 AND PHE-570 DEL; AND VARIANTS HIS-19;		
RP	CYS-54; LYS-238; VAL-259; LYS-400; ARG-575 AND ALA-632.		
RX	MEDLINE-21344600; PubMed-11452359;		
RA	Tissue-Liver;		
RA	Ose L., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,		
RA	Pandya A., Stalenhoef A.F.H., Miettinen T., Bjorkhem I., Bruckert E.,		
RA	Patel S.B.;		
RT	"Two genes that map to the STSL locus cause sitosterolemia: genomic		
RT	structure and spectrum of mutations involving sterolin-1 and		
RT	sterolin-2, encoded by ABCG5 and ABCG8, respectively.";		
RL	Am. J. Hum. Genet. 69:278-290(2001).		
RN	[3]		
RP	REVIEW.		
RA	MEDLINE-2147438; PubMed-11590207;		
RA	Schultz G., Langmann T., Heimerl S.;		
RT	"Role of ABCG1 and other ABCG family members in lipid metabolism.";		
RL	J. Lipid Res. 42:1513-1520(2001).		
CC	- FUNCTION: Transporter that appears to play an indispensable role		
CC	in the selective transport of the dietary cholesterol in and out		
CC	of the enterocytes and in the selective sterol excretion by the		
CC	liver into bile.		
CC	- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to		
CC	excretion.		
CC	- SUBCELLULAR LOCATION: Integral membrane protein (Probable).		
CC	- ALTERNATIVE PRODUCTS:		
CC	Event-Alternative splicing: Named isoforms=2;		
CC	Name=1;		
CC	Isoid-Q9H221-1; Sequence=Displayed;		
CC	Name=2;		

Query Match	Best Local Similarity	99.9%: Score 3502; DB 1; Length 673;
Matches 672; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
FT VARIANT 184	184	R -> H (in sitosterolemia).
FT FT		/Fttd=VAR_012252.
FT VARIANT 231	231	P -> T (in sitosterolemia).
FT FT		/Fttd=VAR_012253.
FT VARIANT 238	238	E -> K.
FT FT		/Fttd=VAR_012254.
FT VARIANT 259	259	A -> V.
FT FT		/Fttd=VAR_012255.
FT VARIANT 263	263	R -> Q (in sitosterolemia).
FT FT		/Fttd=VAR_012256.
FT VARIANT 400	400	T -> K.
FT FT		/Fttd=VAR_012257.
FT VARIANT 405	405	R -> H (in sitosterolemia).
FT FT		/Fttd=VAR_012258.
FT VARIANT 501	501	L -> P (in sitosterolemia).
FT FT		/Fttd=VAR_012259.
FT VARIANT 543	543	R -> S (in sitosterolemia).
FT FT		/Fttd=VAR_012260.
FT VARIANT 570	570	Missing (in sitosterolemia).
FT FT		/Fttd=VAR_012261.
FT VARIANT 572	572	L -> P (in sitosterolemia).
FT FT		/Fttd=VAR_012262.
FT VARIANT 574	574	G -> E (in sitosterolemia).
FT FT		/Fttd=VAR_012263.
FT VARIANT 574	574	G -> R (in sitosterolemia).
FT FT		/Fttd=VAR_012264.
FT VARIANT 575	575	G -> R.
FT FT		/Fttd=VAR_012265.
FT VARIANT 596	596	L -> R (in sitosterolemia).
FT FT		/Fttd=VAR_012266.
FT VARIANT 632	632	V -> A.
FT FT		/Fttd=VAR_012267.
SO SEQUENCE 673 AA; 75678 MW; 594AFDIDCIBB50F CRC64;		
Query Match	99.9%: Score 3502; DB 1; Length 673;	
Best Local Similarity	99.9%: Pred. No. 8.1e-256;	
Matches 672; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
QY 1 MAGAAEBRGLPKGAPDQDTSGLDDRFPSSSEDSNLSFTYSGQNTLEVRDLNAYVDLAS 60		
DB 1 MAGAAEBRGLPKGATPQDTSGLDDRFPSSSEDSNLSFTYSGQNTLEVRDLNAYVDLAS 60		
QY 61 QVPEFDLQAKMWTSPSCNSELGIONLSFYVRSGOMLAIIGSSGCGRASLLDYITG 120		
DB 61 QVPEFDLQAKMWTSPSCNSELGIONLSFYVRSGOMLAIIGSSGCGRASLLDYITG 120		
QY 121 RGHGKIKSGGIWINGOPSSPOLVRKCAVHRQHNOLLPNTVRETAFTAOMLRPTFS 180		
DB 121 RGHGKIKSGGIWINGOPSSPOLVRKCAVHRQHNOLLPNTVRETAFTAOMLRPTFS 180		
QY 121 RGHGKIKSGGIWINGOPSSPOLVRKCAVHRQHNOLLPNTVRETAFTAOMLRPTFS 180		
DB 121 RGHGKIKSGGIWINGOPSSPOLVRKCAVHRQHNOLLPNTVRETAFTAOMLRPTFS 180		
QY 181 QAQRKREVDYIAELRLROCADTRVGNMYVGLSGGERRRYSIGVOLLMPNGILIDEPT 240		
DB 181 QAQRKREVDYIAELRLROCADTRVGNMYVGLSGGERRRYSIGVOLLMPNGILIDEPT 240		
QY 241 SGLDSPFAHNLVKTLRLAKGNRLVLSLHOPKSDIRLDELVLMTSGPIYIGAQM 300		
DB 241 SGLDSPFAHNLVKTLRLAKGNRLVLSLHOPKSDIRLDELVLMTSGPIYIGAQM 300		
QY 301 VQYTAATGYPGCPRTSNADPFYVDLTSIDRSREDELTRREKASIALPLEKVRDLDFL 360		
DB 301 VQYTAATGYPGCPRTSNADPFYVDLTSIDRSREDELTRREKASIALPLEKVRDLDFL 360		
QY 361 WKAETKDLDEDTCESSVTPLDNTCLPSPTKMPGAVOQFTLLIRKQISNDFRDLPTLLIH 420		
DB 361 WKAETKDLDEDTCESSVTPLDNTCLPSPTKMPGAVOQFTLLIRKQISNDFRDLPTLLIH 420		
QY 421 GAERCLMSMTIGLFEHGHSIQLSFMTALLFNIGALLFPNNYLDVYSKCYSERAMLY 480		
DB 421 GAERCLMSMTIGLFEHGHSIQLSFMTALLFNIGALLFPNNYLDVYSKCYSERAMLY 480		
QY 481 ELEEGSLTTPPYFAKILGELPEHCATIIYGMPTVLANRPLQFLLHFLVYLVVF 540		
DB 481 ELEEGSLTTPPYFAKILGELPEHCATIIYGMPTVLANRPLQFLLHFLVYLVVF 540		

OY		541	CRRIMLAAALPPTFHMA	SFSFNALNSYLLAGGMINSSLMATVPAMISKYSFLRMCF	6000
Db		541	CRRIMLAAAAALLPTFHMAS	SFSPNALNSYLLAGGFMINISSLMTPAWMISKYSFLRMCF	6000
OY		601	EGLMKIQFSRRTYKMPDGNLTAV	SGDKILSANELDSYPYLAIYLVIGSGGFMYLYV	6660
Db		601	EGLMKIQFSRRTYKMPDLGNLTAV	SGDKILSWELDSYPLXAIYLVIGSGGFMYLYV	6660
OY		661	SLRFTRKQPSDM	673	
Db		661	SLRFTRKQPSDM	673	
			RESULT 2		
			ABCG8_MOUSE		
ID	ABCG8_MOUSE	STANDARD:	PRT:	673 AA.	
AC	G9DBM0;				
DT	28-FEB-2003 (Rel. 41,				
DT	28-FEB-2003 (Rel. 41,				
DT	28-FEB-2003 (Rel. 41,				
DE	ATP-binding cassette,				
GN	ABCG8.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
ON	NCBT_Taxid=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).				
RC	STRAIN=C57BL/6; TISSUE=Liver;				
RX	MEDLINE=21344600; PubMed=11452359;				
RA	Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,				
RA	Ose I., Stalenhoef A.F.H., Meinen T., Bjorkhem I., Bruckert E.,				
RA	Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,				
RA	Patel S.B.;				
RT	*Two genes that map to the STSL locus cause stoterolemia: genomic				
RT	structure and spectrum of mutations involving sterolin-1 and				
RL	sterolin-2, encoded by ABCG5 and ABCG8, respectively.*;				
RN	Am. J. Hum. Genet. 69:278-290(2001).				
RN	[2]				
RP	SEQUENCE FROM N.A. (ISOFORM 1).				
RC	STRAIN=C57BL/6J; TISSUE=Liver;				
RX	MEDLINE=21083660; PubMed=11217851;				
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,				
RA	Aikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,				
RA	Alkawa K., Ikawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,				
RA	Saito T., Okazaki Y., Gojbori T., Bono H., Kasukawa T., Saito R.,				
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,				
RA	Fletschmann W., Gaasterland T., Gissi C., King B., Kochaya H.,				
RA	Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Mashio T.,				
RA	Sakai K., Okiado T., Furuno M., Anon H., Baldarelli R., Barsh G.,				
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,				
RA	Brownstein M.J., Butt C., Fletcher C., Fujita M., Gariboldi M.,				
RA	Guernicich S., Hill D., Hotmann M., Hume D.A., Kamuya M., Lee N.H.,				
RA	Lions P., Marchionni L., Mashima Y., Mazzarelli J., Mombearts P.,				
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,				
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,				
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming D.,				
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,				
RA	Hayashizaki Y.;				
RT	*Functional annotation of a full-length mouse cDNA collection.*;				
RL	Nature 409:685-690(2001).				
RN	[3]				
RP	TISSUE SPECIFICITY, AND INDUCTION.				
RX	MEDLINE=20553648; PubMed=11099417;				
RA	Berge K.E., Tian H., Graf G.A., Yu L., Grisham N.V., Schultz J.,				
RA	Katerevich P., Shan B., Barnes R., Hobbs H.H.;				
RT	*Accumulation of dietary cholesterol in sitosterolemia caused by				
RT	mutations in adjacent ABC transporters.*;				
RL	Science 290:1771-1775(2000).				
CC	-I- FUNCTION: Transporter that appears to play an indispensable role				
CC	in the selective transport of the dietary cholesterol in and out				

```

CC of the enterocytes and in the selective sterol excretion by the
CC liver into bile.
CC -1 SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to
CC ABCG5 along a pathway regulating dietary sterol absorption and
CC excretion (By similarity).
CC -1 SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1 ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2:
CC Name=1:
CC IsoId=Q9DBM0-1; Sequence=Displayed;
CC Name=2:
CC IsoId=Q9DBM0-2; Sequence=VSP_000053;
CC Note=No experimental confirmation available;
CC -1 TISSUE SPECIFICITY: Expressed in the intestine and, at lower
CC level, in the liver.
CC -1 INDUCTION: Upregulated by cholesterol feeding. Possibly mediated
CC by the liver x receptor/retinoid x receptor (LXR/RXR) pathway.
CC -1 SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG5 (WHITE)
CC SUBFAMILY.
CC -1 CAUTION: Seems to have a defective ATP-binding region.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce,
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AF324495; AAK84079.1; -
CC DR EMBL; AK004871; BAB2630.1; -
CC DR MGD; MGI:1914720; Abcg8.
CC DR InterPro; IPR003439; ABC_transporter.
CC DR Pfam; PF00005; ABC_tran; 1.
CC DR ProDom; PD000006; ABC_transporter; 1.
CC DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
CC DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
CC KW Glycoprotein; Transmembrane; Cytope; Alternative splicing.
CC FT DOMAIN 1 413
CC FT TRANSMEM 414 434
CC FT TRANSMEM 435 447
CC FT TRANSMEM 448 468
CC FT DOMAIN 469 496
CC FT TRANSMEM 497 517
CC FT DOMAIN 518 526
CC FT TRANSMEM 527 547
CC FT TRANSMEM 548 569
CC FT DOMAIN 570 590
CC FT TRANSMEM 591 639
CC FT DOMAIN 640 660
CC FT DOMAIN 661 673
CC FT CARBOHD 619 619
CC FT VARSPLC 377 377
CC FT /FTid=VSP_000053
CC Q9 SEQUENCE 673 AA; 75995 MW; 78012611A5DF2589 CAC64;
CC
CC Query Match 81.9%; Score 2873; DB 1; Length 673;
CC Best Local Similarity 81.8%; Pred. No. 1.7e-208;
CC Matches 551; Conservative 52; Mismatches 69; Indels 2; Gaps 2;
CC
CC QY 1 MAGRAAERGLPKATPDOTS-GLQDLRFSSDSNSLYFTYSGGPNTLEVDLNYVDLA 59
CC DB 1 MAERTKETOLMNTGVTLQDASQGLDLSFSSDSNSLYFTYSGGSNTLEVDLNYVDIA 60
CC
CC 60 SQVWFQFLQAFKMPMTSPSCNSCELGIONLSPKRVSGOMLAIITGSGCGRASILDVYT 119
CC DB 61 SQVWFQFLQAFKLPVMSHSSQDCELGIRLSPKRVSGOMLAIITGSGCGRASILDVYT 120
CC
CC QY 120 GRGHGKTKSGQIININGOPSSPOLVYRCVAVNRQHNOLLPLVYRETLATFAQNRPLPTF 179
CC DB 121 GRGHGKTKSGQIININGOPSSPOLVYRCVAVNRQHNOLLPLVYRETLATFAQNRPLPTF 180
CC
CC 180 SQAQRKRVEDVIALRLQCADTRVGNMTVYGLSGGERRRVSIGVOLLNMPGILIDEP 239

```

DB 181 SOAORRVEDYIAELRLQCANTRGNTYVGVSGGERRRSIGVOLLNRCILLDER 240
QY 240 TSGDSFTAHNVKTLRLAKGNRLVLISLHOPRSDIFRLFDVLMTSGTPIYGAACH 299
DB 241 TSGDSFTAHNVKTLRLAKGNRLVLISLHOPRSDIFRLFDVLMTSGTPIYGAACH 300
QY 300 MVOYFAIGYPCPRNSNPADFYVDLTSDRSREDELAERKAOSIAALEKVRDLDF 359
DB 301 MVOYFAIGYPCPRNSNPADFYVDLTSDRSREDELAERKAOSIAALEKVRDLDF 360
QY 360 LKAKETKDLDEDTCESSVTPDLTNCPSPTKMPGAVOQFTLLIRROISNDRFDLPTLI 419
DB 361 LKAKETKDLDEDTCESSVTPDLTNCPSPTKMPGAVOQFTLLIRROISNDRFDLPTLI 419
QY 420 HGAECALMSMTIGLYFGHGSIQLSFMDTALLFMGALIPNVILDIYSKYSERAMLY 479
DB 420 HGESEALMSMTIGLYFGHGSIQLSFMDTALLFMGALIPNVILDIYSKYSERAMLY 479
QY 480 YLEEGITVGTGYFFAKTIGELPEHCAYIIYGMPTWLANRPGLOPFLHFLVWLV 539
DB 480 YLEEGITVGTGYFFAKTIGELPEHCAYIIYGMPTWLANRPGLOPFLHFLVWLV 539
QY 540 FCCRIMATAAALPLTFPHMASPFSNALYNSFYLAGGFMINTLSIMTVPMWISKVSFLRMC 599
DB 540 FCCRIMATAAALPLTFPHMASPFSNALYNSFYLAGGFMINTLSIMTVPMWISKVSFLRMC 599
QY 600 FEGIMKIOFSRRTYMPMLNTLITAVSGDKILSAMELDXYLAIIYLVIGLSGGFVLY 659
DB 600 FEGIMKIOFSRRTYMPMLNTLITAVSGDKILSAMELDXYLAIIYLVIGLSGGFVLY 659
QY 660 VSLRIRIKORPSDWM 673
DB 660 VSLRIRIKORPSDWM 673
QY 660 LSLKLIRIKORPSDWM 673
DB 660 LSLKLIRIKORPSDWM 673
RESULT 3
ABCG_RAT STANDARD: PRT; 672 AA.
AC P58428;
AT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE ATP-binding cassette, sub-family G, member 8 (sterolin-2).
GN ABCG8.
OS Rattus norvegicus (Rat).
OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN=Sprague-Dawley;
RX MEDLINE=21344600; PubMed=11452359;
RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
RA Ose L., Stalenhoef A.F.H., Miettinen T., Bjorkhem I., Bruckert E.,
RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,
RA Patel S.B.;
RA "Two genes that map to the STSL locus cause sitosterolemia: genomic
RA structure and spectrum of mutations involving sterolin-1 and
RA sterolin-2, encoded by ABCG5 and ABCG8, respectively.";
RL Am. J. Hum. Genet. 69:278-290(2001).
CC -1- FUNCTION: Transporter that appears to play an indispensable role
CC in the selective transport of the dietary cholesterol in and out
CC of the enterocytes and in the selective sterol excretion by the
CC liver into bile.
CC -1- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to
CC ABCG5 along a pathway regulating dietary-sterol absorption and
CC excretion (by similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC Event-Alternative products:
CC Name=1;
CC IsoId=P58428-1; Sequence=displayed;

CC Name=2;
CC IsoId=P58428-2; Sequence=VSP_000054;
CC Note-No experimental confirmation available;
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
CC SUBFAMILY.
CC -1- CAUTION: Seems to have a defective ATP-binding region.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF51785; AKR4831.1; -
CC InterPro: IPR003593; AAA_Atpase.
CC InterPro: IPR003439; ABC_transporter.
CC Pfam: PF00005; ABC_tran; 1.
CC ProDom: PD000006; ABC_transporter; 1.
CC SMART: SM00211; ABC_TRANSPORTER_1; 1.
CC PROSITE: PS00893; ABC_TRANSPORTER_1; 1.
CC PROSITE: PS00893; ABC_TRANSPORTER_2; 1.
CC Glycoprotein; Transmembrane; Transport; Alternative splicing.
CC DOMAIN 1 412
CC DOMAIN 413 433
CC DOMAIN 434 446
CC DOMAIN 447 467
CC DOMAIN 468 495
CC DOMAIN 496 516
CC DOMAIN 517 525
CC DOMAIN 526 546
CC DOMAIN 547 568
CC DOMAIN 569 589
CC DOMAIN 590 628
CC DOMAIN 629 649
CC DOMAIN 650 672
CC CARBOHYD VARSPLIC 376
CC FT 376
CC FT Missing (in isoform 2).
CC FT /FTid=VSP_000054.
CC SQ SEQUENCE 672 AA; 75878 MW; 6088973F3FA4C36F CRC64;
Query Match 80.7%; Score 2829.5; DB 1; Length 672;
Best Local Similarity 79.8%; Pred. No. 3.2e-205;
Matches 537; Conservative 57; Mismatches 78; Indels 1; Gaps 1;
QY 1 MAGKAABERGCLPKGTPDTSGLQDRLTSSSDNSLYFTYSGQPTLEYRDLYVDLAS 60
DB 1 MAGKAABERGCLPKGTPDTSGLQDRLTSSSDNSLYFTYSGQPTLEYRDLYVDLAS 60
QY 1 MAQTTKEFTQLMNGTVLQDASSLQDSVFSSESDNSLYFTYSGQPTLEYRDLYVDLAS 60
DB 1 MAQTTKEFTQLMNGTVLQDASSLQDSVFSSESDNSLYFTYSGQPTLEYRDLYVDLAS 60
QY 61 QVPMFEOLAQFKMPTSPSCONSCGLGIONISFYKRSQOMLITSSGGRASLDVDTG 120
DB 61 QVPMFEOLAQFKMPTSPSCONSCGLGIONISFYKRSQOMLITSSGGRASLDVDTG 120
QY 61 QVPMFEOLAQFKMPTSPSCONSCGLGIONISFYKRSQOMLITSSGGRASLDVDTG 120
DB 61 QVPMFEOLAQFKMPTSPSCONSCGLGIONISFYKRSQOMLITSSGGRASLDVDTG 120
QY 121 RGHGKIKSGQIWMINGOPSSPOLYKCAVHVRQHNQLPNTLVRETLAFIAQMRLPRTFS 180
DB 121 RGHGKIKSGQIWMINGOPSSPOLYKCAVHVRQHNQLPNTLVRETLAFIAQMRLPRTFS 180
QY 121 RDHGGKMKSGQIWMINGOPSSPOLYKCAVHVRQHNQLPNTLVRETLAFIAQMRLPRTFS 180
DB 121 RDHGGKMKSGQIWMINGOPSSPOLYKCAVHVRQHNQLPNTLVRETLAFIAQMRLPRTFS 180
QY 181 QAOQRKREVEDYIAELRLQCANTRGNTYVGVSGGERRRSIGVOLLNRCILLDEPT 240
DB 181 QAOQRKREVEDYIAELRLQCANTRGNTYVGVSGGERRRSIGVOLLNRCILLDEPT 240
QY 241 SGDSFTAHNVKTLRLAKGNRLVLISLHOPRSDIFRLFDVLMTSGTPIYGAACH 300
DB 241 SGDSFTAHNVKTLRLAKGNRLVLISLHOPRSDIFRLFDVLMTSGTPIYGAACH 300
QY 301 VQYFTSISYPCPRNSNPADFYVDLTSDRSREDELAERKAOSIAALEKVRDLDF 360
DB 301 VQYFTSISYPCPRNSNPADFYVDLTSDRSREDELAERKAOSIAALEKVRDLDF 360
QY 361 WKAETKDLDEDTCESSVTPDLTNCPSPTKMPGAVOQFTLLIRROISNDRFDLPTLI 420
DB 361 WKAETKDLDEDTCESSVTPDLTNCPSPTKMPGAVOQFTLLIRROISNDRFDLPTLI 420

OY 421 GAELACMSMTIGELFYHGSIOLSEFMDTALLFMGALIPFNVLIDYISKYSERAMLYX 480
 DB 420 GAELACMSMTIGELFYHGSIOLSEFMDTALLFMGALIPFNVLIDYISKYSERAMLYX 479
 OY 481 ELEDGLYTPGPPFAKTLGELPEHCAYIIYGMFTYLANLRPGLOPFLHFLHVLVWF 540
 DB 480 ELEDGLYTPGPPFAKTLGELPEHCAYIIYGMFTYLANLRPGLOPFLHFLHVLVWF 539
 OY 541 CCRIMALAAALPTFMASFFSNALYNSFYLAGCFEMINISLMTVAANISKYSFLMWF 600
 DB 540 CCRIMALAAALPTFMASFFSNALYNSFYLAGCFEMINISLMTVAANISKYSFLMWF 599
 OY 601 EGLMKIOFSRRTYKMPGLNLTIAVSGDKILSAMELDSPYATYLYIGLSGFMVLYV 660
 DB 600 SGLMOIOFNGHIITTOIGNLTFVYPGDAMVTAMDLSHPLXATYLYIGLSGFLSLYL 659
 OY 661 SLAFIKOPSQDW 673
 DB 660 SLAFIKOPSQDW 672

RESULT 4

ABG5_RAT STANDARD; PRT; 652 AA.
 ID ABG5_RAT 099PE7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 5 (Sterolin-1).
 GN ABCG5.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId=10116;
 RN [1]
 RA Dean M. Patel S.B.;
 RA Allketers R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,
 RT "Identification of a gene, ABCG5, important in the regulation of dietary cholesterol absorption.";
 RL Nat. Genet. 27:79-83(2001).
 CC -1- FUNCTION: Transporter that appears to play an indispensable role in the selective transport of the dietary cholesterol in and out of the enterocytes and in the selective sterol excretion by the liver into bile.
 CC -1- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to ABCG8 along a pathway regulating dietary-sterol absorption and excretion (by similarity).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE) SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC EMBL: AF317174; AAC53098.2;
 DR InterPro: IPR003593; AAA_ATPase.
 DR Pfam: PF00005; ABC_tran.1.
 DR ProDom: PD000006; ABC_transporter.1.
 DR SMART: SM00382; AAA.1.
 DR PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE: PS00893; ABC_TRANSPORTER_2; 1.
 KW ATP-binding; Glycoprotein; Transmembrane; Transport.

FT DOMAIN 1 385 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 386 406 1 (POTENTIAL).
 FT DOMAIN 407 422 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 423 443 2 (POTENTIAL).
 FT DOMAIN 444 463 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 464 484 3 (POTENTIAL).
 FT DOMAIN 485 504 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 505 525 4 (POTENTIAL).
 FT DOMAIN 526 529 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 530 550 5 (POTENTIAL).
 FT DOMAIN 551 624 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 625 645 6 (POTENTIAL).
 FT DOMAIN 646 652 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 652 652 ATP (POTENTIAL).
 FT CARBOHYD 87 94 N-LINKED (GLCNAC: . .) (POTENTIAL).
 FT CARBOHYD 585 585 N-LINKED (GLCNAC: . .) (POTENTIAL).
 FT CARBOHYD 592 592 N-LINKED (GLCNAC: . .) (POTENTIAL).
 SQ SEQUENCE 652 AA; 73342 MW; 4D42FE2BABDDAD59 CRC64;

Query Match 20.38; Score 713; DB 1; Length 652;
 Best Local Similarity 30.08; Pred. No. 5.6e-46;
 Matches 190; Conservative 115; Mismatches 232; Indels 96; Gaps 15;

OY 12 PKGAT-PODTSGLQDRLEFSSEDSNLSLFTYSGQNTLEVRDNLQVQLASQV-PWFQOLA 69
 DB 9 PEGARGHNNRGSO-----SSLEBSGV--TSEARHSISGV--LVNFSVSNNRGVW----- 55
 OY 70 QFKMPWTPSPSCNSCELG-I-QNLSFKYKSGOMALITSSCGRASLSDVITGRGKGK 128
 DB 56 -----WNIKSCQCKWKDKILKDVSLYESQOTICILSSSGKTTLLDALSGLRRRTGL 110
 OY 129 SGLMINGQPSQOLVARKCAVYRQNNQLPNTVRETTAFIAQMLRPTFSQAQRDRV 188
 DB 111 EGEVFNVCGLRBDQFQDCVSYLQSVFSLSTVRETTLYTAMAL-RSSADFYDKV 169
 OY 189 EDVIAELRLQCADTRGNMYVRGLSGERRRYSIGQLMNGIILIDPESGLDSFTA 248
 DB 170 EAVTLTSLSHVADQMGNTGNFGISGERRRYSIAQLQDPRVMMLDPTTGLDCMTA 229
 OY 249 HNLVKTLSRLAKNRLVLSLHOPRSDIFRFLVLLMTSGTPIYLGAQAHVQYFPAIG 308
 DB 230 NHIVLLVLELARNRRIYIVTHOPRSELHFHFKIALITYGELVPGCTPREMGFFRNCG 289
 OY 309 YPCPRNSNPRDFVDLSDIRSRQDLATREKAQSLAIF-----LEKVRDL 356
 DB 290 YPCPRNSNPRDFVDLSDIRSRQDLATREKAQSLAIF-----LEKVRDL 349
 OY 357 DDFLMKAEFTKDLDEDCEVSSVPLDTNCLSPSTK-MPGAVQFTLLIRQISNDPRDL 415
 DB 350 -----KTLPM-----VPEFTKNPGEFKLGVLLRVTNRNLRNG 385
 OY 416 TLLIHGAELCMSMTIGF--LYFGHSGIOLSEFMDTALLFMGALIPFNVLIDYISKYS 473
 DB 386 VVIMRLVQVNLIMGLFLIFLTLRYQNNMKGAVQDRGLXQLVGARPYTGMLNAVLPFM 445
 OY 474 ERAMLTYELEDGLYTPGPPFAKTLGELPEHCAYIIYGMFTYLANLRPGLOPFLHFL 533
 DB 446 LRAVSDQESODIGLYOKRMOMLAVLHALPFSIYAVIFSSVCWTGLGVPEVARF----- 500
 OY 534 LVWLVECCRIMALAAALPTFMASFFSNAL-----INSFLAG 575
 DB 501 -----GYSAALLPRLIGELFTLVLLGVQNPNTVNSIVALLSISGLIGSG 548
 OY 576 FMINLSLMTVPAMISVFLKCFGLMKIOR 608
 DB 549 FIRNIEMPIPLKILGYFTFFOKYCCILVYNEF 581

RESULT 5

ABG5_HUMAN STANDARD; PRT; 651 AA.
 ID ABG5_HUMAN 09H222;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)

[illegible]

```

DB 295 SNPEFDYMDLTVDTSKERELETSTSRVOMIESAVKRSKAICHTKLNIEEMKHL----- 348
OY 363 AETKDEDEDCVSSVTPDNTCLPSPTR-MPGAQOFTTLIRQROSNDFRDLPTLLING 421
DB 349 -----KLPW-----VPEKTDSPGVSKLGVLLRRVNRNKLAVITRL 390
OY 422 ABAACMSMTIGFLYFC-----HGSIOLEMDTALLFIMGALIPNVILDIVSKYSR 475
DB 391 LQNLIMGLLEFLFVLRVSNVAKGAIO-----DRVGLLYFGVAGPTTGKLNANLEPVLR 446
OY 476 ALAYLEDEGLYTTGTYFFAKILGELPEHCATIIYGMPTVLANIPGLPFLHFLV 535
DB 447 AVSDDESODGLYOKQOMLALVALHVLPFSVATMIFSSVYWTGLGHPVARR----- 499
OY 536 WLVEFCRIMALAALPTFMHASFPS-----NALYMSFLAG-----GFM 577
DB 500 -----GYFSALLAPHLIGELTLVLGIYONPNIVNSVALLISAGVLGSGPL 549
OY 578 IULSSIMTVPANISKVSFLRMCPEGIMKIQFSRRTYKMPDGLNLTIVAS 625
DB 550 RNIOEMPDPKIIISYFTFOKYSCEILVNAEFGLNFTCGSSNVSVTTN 597

RESULT 6
ABG5_MOUSE
ID ABG5_MOUSE STANDARD; PRT; 652 AA.
AC 099PE8;
DT 28-FEB-2003 (Rel. 41, Last Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 5 (Stereoloin-1).
GN ABCG5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Liver;
RX MEDLINE=20578753; PubMed=11138003;
RA Lee M.-H., Lu K., Hazard S., Yu R., Shulenin S., Hidaka H., Kojima H.,
RA Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,
RA Dean M., Patel S.B.;
RT Identification of a gene, ABCG5, important in the regulation of
RT dietary cholesterol absorption.;
RL Nat. Genet. 27:79-83(2001).
RN 12
RP TISSUE SPECIFICITY, AND INDUCTION.
RX MEDLINE=20553648; PubMed=11099417;
RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
RA Kvitlerovich P., Shan B., Barnes R., Hobbs H.H.;
RT Accumulation of dietary cholesterol in sitosterolemia caused by
RT mutations in adjacent ABC transporters.;
RL Science 290:1771-1775(2000).
CC -1- FUNCTION: Transporter that appears to play an indispensable role
CC in the selective transport of the dietary cholesterol in and out
CC of the enterocytes and in the selective sterol excretion by the
CC liver into bile.
CC -1- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
CC ABCG8 along a pathway regulating dietary-sterol absorption and
CC excretion (by similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC level, in the liver.
CC -1- INDUCTION: Upregulated by cholesterol feeding. Possibly mediated
CC by the liver X receptor/reinoic X receptor (LXR/RXR) pathway.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL: AF312713; AMG53097.1; -
DB MGD: MG1:1351659; Abcg5.
DB InterPro: IPR003593; AAA_ATPase.
DB InterPro: IPR003439; ABC_transporter.
DB Pfam: PF00005; ABC_tran; 1.
DB PRODOM: PD000006; ABC_transporter; 1.
DB SMART: SM00382; AAA; 1.
DB PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
DB PROSITE: PS00893; ABC_TRANSPORTER_2; 1.
DB ATP-binding: Glycoprotein; Transmembrane; Transport.
DB DOMAIN 1 (POTENTIAL).
DB TRANSMEM 386 406
DB DOMAIN 407 422
DB TRANSMEM 423 443
DB DOMAIN 444 463
DB TRANSMEM 464 484
DB DOMAIN 485 504
DB TRANSMEM 505 525
DB DOMAIN 526 529
DB TRANSMEM 530 550
DB DOMAIN 551 622
DB TRANSMEM 623 643
DB DOMAIN 644 652
DB NP_BIND 87 94
DB CARBOHYD 410 410
DB CARBOHYD 585 585
DB CARBOHYD 592 592
DB SEQUENCE 652 AA; 73244 MW; 80CE37ADCC19771E CRC64;

Query Match 19.7%; Score 691.5; DB 1; Length 652;
Best Local Similarity 28.6%; Pred No. 2.3e-44;
Matches 188; Conservative 129; Mismatches 241; Indels 99; Gaps 16;

OY 45 NLEVRDNLNYVDLASOV-PWEQLAOFKMPYSPSCONSCELGI-QNLSPFVRSQOMLA 102
DB 37 HSLGLVHVSYSV--SNRGPW-----WNISCCQKMRQLKDVSLIESQINC 84
OY 103 IIGSSGCRASLDVITGRHGCKIKSQIWINQPOSSQVLVKKCAVYARQNNQLPNT 162
DB 85 ILGSSGSKRTLLDLSICRLRTGLCEVEFVNGCELRRDPQDFSVYLGSDVFLSLT 144
OY 163 VRETLAFTAOHRLPTFSQAQDRKREVIKELRGQADRTVGMVYRGLSGERRRVS 222
DB 145 VRETLRYTAMALCKS-SADTYNKVEAVWELSHVADQIGSYNFGISGERRRVS 203
OY 223 IGVOLLNMPGILILDEPTSGDSTAHNLVYTLISLAKGNRLVTLISLQPRSDIRPLDL 282
DB 204 IAAQLLOPRTVMIDEPPTGDDCMANQIVLLAELARDRVITYTIHQPSRELFQHQDK 263
OY 283 VLMTSGPIYVLGAQHVQYFTALGYCPKRSNPADYVDLTSTIDRSRQDELATRKA 342
DB 264 IAILTYGLVFCGPPEEMIGFNNCGYCPCHSNPFEDYMDLTVDTSQSREREITYRV 323
OY 343 QSLAALFEKVRDLDDELKMAKETDLDDEDCVSSVTPDNTCLPSPTR-MPGAQOFTT 401
DB 324 OMLECAFE-----SDIYHKI-LENTIEARLYKLTPT-----VPEKTDSPGMGKLG 371
OY 402 LIRQISNDFRDLPTLLIHGAELMSMTIGF--LYFGHSIOLEMDTALLFIMGALI 459
DB 372 LLRRVTRLMKRNKQAVIMRLVQNLIMGLFELFYLRVQNNLTAKGAVQDRVLLIOLVAT 431
OY 460 PENVILDIVSKYSERAMLYLEDEGLYTTGTYFFAKILGELPEHCATIIYGMPTVMA 519
DB 432 PYTGMLNAVNLFPMLRAVNSDQDGLYHKQOMLALVALHVLPFSVATMIFSSVYWT 491
OY 520 NLREGLOPFLHFLVNLVYVCCRIMALAALPTFMHASFNSAL----- 566
DB 492 GLIPEVARF-----GYFSALLAPHLIGELTLVLGIYONPNIVNSI 534

```

OY 567 -----YNSFYIAGGFMINLSLMTVPAMISKVSFLMCEGLKIOFSRRTYMPJLONT 621
 DB 535 VALLSTISGLISGIFRMIQEMPIPKLIIFFYFORXCEIIVANFE-----YGL-----NFT 587
 OY 622 IIVSGDKILSAMELSDSYPLVAI-----YLVIGISGFMVL 657
 DB 588 CGGSNMSML-----NHPMCAITGCVQIFIEKTCPCGATSRFTANFLILGTPALVIL 638
 RESULT 7
 WHIT. DROME STANDARD; PRT: 687 AA.
 ID WHIT. DROME STANDARD; PRT: 687 AA.
 AC P10090: G9Y3A2: G9Y33; 687 AA.
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE White protein.
 GN W OR EG: BACN33B1.1 OR: CG2759.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Head;
 RA MEDLINE=90221897; PubMed=2109311;
 RA Peping M., Mount S.M.;
 RT *Sequence of a cDNA from the *Drosophila melanogaster* white gene.;
 RL Nucleic Acids Res. 18:1633-1633(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85134865; PubMed=6084717;
 RA O'Hare K., Murphy C., Lewis R., Rubin G.M.;
 RT *DNA sequence of the white locus of *Drosophila melanogaster*.;
 RL J. Mol. Biol. 180:437-455(1984).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21100348; PubMed=11156992;
 RA Lukacsovich T., Asztalos Z., Awano W., Baba K., Kondo S., Niva S.,
 RA Yamamoto D.;
 RT *Dual-tagging gene trap of novel genes in *Drosophila melanogaster*.;
 RL Genetics 157:727-742(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Gelniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.C., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Abmayyan A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.T., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Bortova K., Boudreau P., Boudreau P., Boudreau P., Boudreau P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dopp L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostel C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heilmann T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jalaali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh N.V., McLeod M.P., McPherson D.,
 RA Merulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Massaman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT *The genome sequence of *Drosophila melanogaster*.;
 RL Science 287:2185-2195(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R;
 RX MEDLINE=20196011; PubMed=10731137;
 RA Benos P.V., Galt M.K., Ashburner M., Murphy L., Harris D.,
 RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demillies J., Cadieu E.,
 RA Dreano S., Gloux S., Lelaure V., Mottier S., Gallibert F., Borkova D.,
 RA Minna B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,
 RA Papanikolaou G., Spanos L., Cox S., Madueno E., de Pablos B.,
 RA Modolell J., Peter A., Schaefer U., Jaekle H., Mouriotti F.,
 RA Belnert N., Dowe G., Schaefer U., Jaekle H., Mouriotti F.,
 RA Callister D.M., Campbell L.A., Darlamitson A., Henderson N.S.,
 RA McMillan P.J., Sallies C., Tait E.A., Valenti P., Saunders R.D.C.,
 RA Glover D.M.;
 RT *From sequence to chromosome: the tip of the X chromosome of *D. melanogaster*.;
 RL Science 287:2220-2222(2000).
 RN [6]
 RP SEQUENCE OF 224-331 FROM N.A.
 RX MEDLINE=89339145; PubMed=2503416;
 RA Teatle R.G., Belote J.M., McKewen M., Baker B.S., Howells A.J.;
 RT *Cloning and characterization of the scarlet gene of *Drosophila melanogaster*.;
 RL Genetics 122:595-606(1989).
 CC -1- FUNCTION: PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT CELLS RESPONSIBLE FOR EYE COLOR. WHITE DIMERIZE WITH BROWN FOR THE TRANSPORT OF GUANINE AND WITH SCARLET FOR THE TRANSPORT OF TRYPTOPHAN.
 CC -1- SUBUNIT: HETERODIMER OF WHITE WITH EITHER BROWN OR SCARLET.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@sdb-sdb.ch).
 CC -----
 DR EMBL: X51749; CAA36038.1;
 DR EMBL: X02974; CAA26716.1;
 DR EMBL: AB028139; BAA78210.1;
 DR EMBL: AE003425; AAR45826.1;
 DR EMBL: AL133506; CAB65847.1;
 DR EMBL: X76202; CAA53795.1;
 DR PIR: S08635; FYFPM.
 DR FlyBase: FBgn0003996; w.
 DR GO: GO:0004888; P:transmembrane receptor activity; NMS.
 DR GO: GO:0006727; P:chromosome biosynthesis; IMP.
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transporter.
 DR InterPro: IPR005284; Pigment_permease.
 DR Pfam: PF00005; ABC_tran.1.
 DR ProDom: PD000006; ABC_transporter.1.
 DR SMART: SM00382; AAA.1.
 DR TIGRFAMS: TIGR00953; 3401204.1.
 DR PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE: PS00893; ABC_TRANSPORTER_2; 1.

KM Placement: ATP-binding; Transmembrane; Transport.
 FT NP_BIND 130 137 ATP (BY SIMILARITY).
 FT TRANSMEM 435 453 POTENTIAL.
 FT TRANSMEM 465 485 POTENTIAL.
 FT TRANSMEM 515 533 POTENTIAL.
 FT TRANSMEM 542 563 POTENTIAL.
 FT TRANSMEM 576 594 POTENTIAL.
 FT TRANSMEM 659 678 POTENTIAL.
 FT CONFLICT 25 29 GDSGA -> LIFEIPHYCHRYAD (IN REF. 2 AND 3).
 FT CONFLICT 49 49 L -> R (IN REF. 4 AND 5).
 FT CONFLICT 335 371 VGACCPNYPNADYOVYLAIVVGPGEIESRDRKAKIC -> ITHLNSYPAWVSPVSLPTTIRPFTYTRCMPLCPGRCSSPVI GSPRYG (IN REF. 3).
 FT SEQUENCE 687 AA; 75672 MW; 24AFAD799DEDD36 CRC64;
 Query Match 18.7%; Score 656; DB 1; Length 687;
 Best Local Similarity 30.3%; Pred. No. 1,2e-41;
 Matches 178; Conservative 113; Mismatches 265; Indels 32; Gaps 10;
 OY 88 IONLSFKVRSQGMALITGSSGGRASLDVITGRGHG--KISGQIWMNGPSSPOLVR 145
 DB 113 LKAVCGVAIPGELLAVWSSGAGKRTLLNALFRSPGIGVSPSGMRLNGOPADAKEMQ 172
 OY 146 KCVAHYRHNOLPNTVRETLAFIAQMRIPFESQQRDRVEDVIAELRQCADTRV 205
 DB 173 ARCAVYQODOLFISLTAHREHLIFQANVRPRHLITQRARVDQVIELSKCHTTI 232
 OY 206 G-NMYRGISGGRRRVSTIGVOLLNMGILLDEPTSGDSTAHNLVKTLSRLAKNRL 264
 DB 233 GVPGRVKGISGGRKRRLAFSEALTPDLICDEPTSGDSTAHNSVYVYKLSKSGKT 292
 OY 265 VILSLHOPSDIRFLDVLVLTMSGTPPIYLGAAQHWQYTAIGYPCPRSNADRYVDL 324
 DB 293 VITTIHOPSELELTDKILMAEGRVAFIATPSEADFSYGAQCPNYPNADRYVQV 352
 OY 325 TSIDRRSRDELATREKAOSLALF-LEKY-RDLDFLMAETKDLDEDTCESSVYTPK 382
 DB 353 LNV---VPGREIESRRIKICDNFAISKVARMEDL---ATKNLEK-----PLE 397
 OY 383 TNCLEPSP---TKMCAVOOFTLLIRQISNDRFLPTLLIHGAECNLSMTIGTFYFGH 438
 DB 398 ---OPENGYTKATWFOFRAVLWRSMLSVLKEPLVAVRLIQTWVAIILGLLEFGQ 452
 OY 439 GSISLSPMDALFEMGALIPPNVLDVYSKCYSRAMLYLEGLTITGTFYPAKIL 498
 DB 453 QLTQGVMMINGAIFLELTNMTFQNFATINFTSELPEFMRARSRLYRCQDYFFLGKTI 512
 OY 499 GELPEHCAYIIYGMPTWLANLRPGLOPELLAFLLVMTLVVFCRIMATAAALLPTEHM 558
 DB 513 AELPELFTVPIVFTALAYMIGRACVLFENCLALVTVLVANSTFGILISCASSSTMS 572
 OY 559 ASFEFSAALYNSETLAGFMINLSLTWVAMISKVSFLRMCFEGILKTIQFS---RRTYKA 615
 DB 573 ALSVGPVYIIPLELFGFPLNSGVPVYKMLSYLSMFRANGLINQMDADVEPEISG 632
 OY 616 PLGNLTINVSQDKIISAMELSDYPLATVILVIGLSGGWVLYVSLR 663
 DB 633 TSSNTCPSSGKVIETLNFNSADPLDVLVGLAIIIVSPRVILAYLR 680
 RESULT 8
 YOH5_YEAST STANDARD; PRT: 1294 AA.
 AC 008234; 008233;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable ATP-dependent transporter YOL074C/YOL075C.
 GN YOL074C/YOL075C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97321807; PubMed=9178509;
 RA Tzeemia M., Katsoulou C., Alexandraki D.;
 RT "Sequence analysis of a 33.2 kb segment from the left arm of yeast chromosome XV reveals eight known genes and ten new open reading frames including homologues of ABC transporters, Inositol phosphatases and human expressed sequence tags.";
 RT Yeast 13:583-589(1997).
 RL
 CC 1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC 1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@sdb-sdb.ch).
 CC
 CC EMBL: Z74817; CAAG9085.1; -
 CC EMBL: Z74816; CAAG9084.1; -
 CC PIR: S77690; S77690.
 CC SGD: S0005435; YOL075C.
 CC InterPro: IPR003593; AAA_ATPase.
 CC InterPro: IPR003439; ABC_transporter.
 CC Pfam: PF00005; ABC_tran; 2.
 CC Prodom: PD000006; ABC_transporter; 2.
 CC SMART: SM00382; AAA; 2.
 CC PROSITE: PS00211; ABC_TRANSPORTER_1; 2.
 CC PROSITE: PS00893; ABC_TRANSPORTER_2; 2.
 CC Hypothetical protein; ATP-binding; Transmembrane; Glycoprotein;
 KW Transport; Repeat.
 KM
 FT TRANSMEM 376 396 POTENTIAL.
 FT TRANSMEM 496 516 POTENTIAL.
 FT TRANSMEM 531 551 POTENTIAL.
 FT TRANSMEM 605 625 POTENTIAL.
 FT TRANSMEM 1039 1059 POTENTIAL.
 FT TRANSMEM 1121 1141 POTENTIAL.
 FT TRANSMEM 1267 1287 POTENTIAL.
 FT NP_BIND 62 69 ATP (POTENTIAL).
 FT NP_BIND 727 734 ATP (POTENTIAL).
 FT CARBOHYD 41 41 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 86 86 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 101 101 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 151 151 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 341 341 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 349 349 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 371 371 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 528 528 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 983 983 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1062 1062 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 1294 AA; 145157 MW; C555500A5B5284E CRC64;
 Query Match 18.6%; Score 653; DB 1; Length 1294;
 Best Local Similarity 30.1%; Pred. No. 4.4e-41;
 Matches 171; Conservative 111; Mismatches 239; Indels 48; Gaps 13;
 OY 88 IONLSFKVRSQGMALITGSSGGRASLDVITGRGHGKIKSQI----- 132
 DB 45 VNTFSMDLPSSGSMAMVGGSGKTTLLNVLASKISGGLHNSIRYVLEDGSEPNETE 104
 OY 133 ---WINGPSSPOLVARKCVANRHNOLPNTVRETLAFIAQMRIPFESQQRDR- 187
 DB 105 PKRAHLDGQ-DHPYQKHVIMATVLPQDVLSPRLTCHETLKFADLKL-----NSSERTKL 159
 OY 188 -VEDVIAELRLQCADTRVGNMYVRLSGGERRRVSIGVOLLNMGILLDEPTSGIDSP 246
 DB 160 WVEQLLEELGLKCAQDVLGDNDSHRLSGEKKRRLSIGVOMSNPSIMFLDEPTGLDAV 219
 OY 247 TANHVLKTLRLAK-GNRLVLSLHOPSDIRFLDVLVLTMSGTPPIYLGAAQHWQYTP 305

FT CONFLICT 208 208 F -> S (IN REF. 1).
 FT CONFLICT 315 316 MISSING (IN REF. 5).
 FT CONFLICT 482 482 R -> T (IN REF. 2).
 SO SEQUENCE 655 AA; 72343 MW; 894603511DC5C60 CRC64;

Query Match 18.3%; Score 640.5; DB 1; Length 655;
 Best Local Similarity 27.9%; Pred. No. 1.6e-40;
 Matches 175; Conservative 131; Mismatches 254; Indels 67; Gaps 17;

OY 80 CONSELGI-QNLKRVSGOMLAIGSSGGRSLVYITGRGKIGKQWINGOP 138
 DB 55 CRKPEKEILSNKIMKPG-LMALGPTGGKSSLDVLAARKDPSGL-SGDVILINAP 112
 OY 139 SSPOLVRKC-VAHROHNOQLPNTVRETIAFIAQMLPPEFSOAKORVEDVIAEURL 197
 DB 113 RPANF-KCNSGYVODVVGTLVRNLOPFAALRLATWTNHEKMERINRYEEGL 170
 OY 198 RQCADTRYGNMYVGLSGERRRVSIGVOLLMPGILLDEPTSGDSFTAHNLVKTLSR 257
 DB 171 DKVADSKVGTQPIGVSIGERRKRTSIGMELITDPSILFDEPTGIDSTANAVALLKLR 230
 OY 258 LAKGNRLVLSLHOPRSDIEPLDVLMTSGTPIYGAQAHMOYFAIGPCPRYSNP 317
 DB 231 HSKGRITLIESIHOPRSTIFKLPDLSLTLASGRIMFHGPAQDALGYEFSAGYHEAYNNP 290
 OY 318 ADPYVDLTSIDRR-SREOLATRE-KAOSLAALFLEKVRDL-DDFLMKAKETK--- 366
 DB 291 ADPLDILINDSTVALNREDFKATELIEPSKODKPLEIAETIYVNSSTFK-ETKAE 349
 OY 367 -----DLDEPTCVESVTPIDNCLPSPTKPGAVOQFTLLIROISNDFRDLPTLLIH 420
 DB 350 HOLSGGKRRKITYFKKISYTSFSC-----HQLRVNSRSEFNLLGNQASIAQ 398
 OY 421 GAELCKMTGILYFEGHSIQLSFMDTALFEMIGALIPNVLDVSKYS----- 473
 DB 399 IYTVVGLVIGALYFGLKNDSTGIONRAGYLFPL-----TTNOCFSSVSAVEL 447
 OY 474 ---ERAMLYLEEDGLTYTGPYFPAKILGE-LPEHCAYIIITGMPTWLANLRGIDPFL 529
 DB 448 FVEEKKLFIEHYISGYRVSSYFGLKSLDLPRMLPSITFCITVFMGLKRADEF 507
 OY 530 LHELVLVLFVCCRIMALAAALPTEPHMASFEFSNALYNSFYLAGCFMILNLSMTYDAM 589
 DB 508 VMETLMAVAYASASMALATAAGOSVSVATILMTFCVFEMIMISGLLVNTITASHLSW 567
 OY 590 ISKVSFLRMCFEGMLKQFSRRTYKMLGULT-----IAVSDKTL--SAMELDSYPL 640
 DB 568 LQYFSTIPRYGFTALQHNHFLGONF-CPLNATGNNPCNVAATCTGEYLVKQIDLSPMGL 626
 OY 641 YAIYLVIGLGGFMVLYVSLRFIKO 667
 DB 627 WKNHVALACMIVIFLTITAIKLFLFK 653

RESULT 10
 WHIT-ANOGA STANDARD; PRT; 695 AA.
 AC 027256: 017006;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE White protein.
 GN M.
 OS Anopheles gambiae (African malaria mosquito).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
 OX NCBI_TaxID=7165;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPRATN-Snakoko / G3;
 RX MEDLINE=96423158; PubMed=8825759;
 RA Besansky N.J., Bedell J.A., Benedict M.O., Mukabayire O., Hliffiker D., Collins F.H.;

FT Cloning and characterization of the white gene from Anopheles
 FT gambiae.
 RL Insect Mol. Biol. 4:217-231(1995).
 CC -1- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
 CC NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
 CC CELLS RESPONSIBLE FOR EYE COLOR.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U29486; AAC46995.1; -
 DR EMBL: U29485; AAC46994.1; -
 DR EMBL: U29484; AAC47423.1; -
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transporter.
 DR InterPro: IPR005284; Pigment-permease.
 DR Pfam: PF00005; ABC_tran.1.
 DR Prodom: PD00006; ABC_transporter.1.
 DR SMART: SM00382; AAA.1.
 DR TIGRfam: TIGR00955; 3a01204.1.
 DR PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE: PS00893; ABC_TRANSPORTER_2; 1.
 KW Pigment; ATP-binding; Transmembrane; Transport.
 FT NP_BIND 133 140
 FT NP_BIND 288 295
 FT TRANSMEM 444 464
 FT TRANSMEM 474 494
 FT TRANSMEM 524 544
 FT TRANSMEM 552 572
 FT TRANSMEM 581 601
 FT TRANSMEM 669 689
 FT CARBOHYD 472 472
 FT CARBOHYD 645 645
 FT CONFLICT 100 100
 FT CONFLICT 691 693
 SO SEQUENCE 695 AA; 77218 MW; EB8B951723982961 CRC64;

Query Match 17.9%; Score 627; DB 1; Length 695;
 Best Local Similarity 26.3%; Pred. No. 1.8e-39;
 Matches 189; Conservative 128; Mismatches 289; Indels 112; Gaps 17;

OY 14 GATPDQISGLDRLFFSES-----DNSLYFTYSGOPNTL-EVRDLNLYVDLASQVPEF 66
 DB 10 GDAESKTTISSRRYSSSYODSDMDALNTLLNDRKATLIQVWKPKSYGVKQIPCE 69
 OY 67 QLAQFKMPW-----TSPSC--ONSCELG-----IONLSFKYRSG 98
 DB 70 RLTF---YTWEIDYVFGCAPPDGKREPLCTRLNCCYRORRDNPRKHLNVTGVAKSG 126
 OY 99 QMLAIIGSSGGRASLDVITGRGHC-KIKSGOI-WINQPSSPQVLARKVAHVRQNO 156
 DB 127 ELLAVMGSSGAGKTTLLNALAFRSPGVKISPNVRLNCGVPAEQLRARCAYVQDD 186
 OY 157 LRLPLTYRETLAFIAQRRLPRTSQAQRKVEDVIAELRLRQCADRVGNN-YVRGLSG 215
 DB 187 FIPSLTRREHLIDPAMLRMGDRDPAVSKVORVQEVQLDELVLVCAQDTITGAPRIKGLSG 246
 OY 216 GERRRVISIGVOLLMPGILLDEPTSGDSFTAHNLVKTLSRLAKGNRLVLSLHOPRSD 275
 DB 247 GERKRLAFASSETLIDPULLLDEPTSGDSFMASVADVLKGAAMKRTITLTIHOPSS 306
 OY 276 IFRLEDLVLLMTSGTPIYGAQAHMOYFAIGPCPRYSNPADFYVDLTSIDRRSRE 335
 DB 307 LYGCFDLILLVAGBRVAFSSPQSAEFFSGDLIPCPRNPNPDEYQMLAIAPAK----- 362
 OY 336 LATREKQSLAALFLEKVRDLDFLWKAETKDLDEPTCVESVTPIDTNCLPSPT----- 390


```

Db 363 -----EACRDMIRKICDSFVSPISAREVLETAASVAGKG 396
OY 391 -KMPGAVQ-----OFTLLIRROIENDPRDLPTLLIHGAELMTGTF 433
Db 397 MDEPYMLOOEGCVGTGRSSWMTQFYCLIMRSMLSYKDPMLVAVRLLOTAAMAVTLIGS 456
OY 434 LYFGHGSIOLEPMPTAALLFMIGALIPFNVLIDVTSKCYSERAMLYELEDGLYTGPFY 493
Db 457 IYFGVLDLQDDGMVNMINGSLFELTMMTFQNVFAVINVSAPLEPVLFRKRSRLYVVDYTF 516
OY 494 FAKILGELPEHCAYIIIGMPYWMANLRPGLOPPLFLVLMVLFECRRIMAAAL 553
Db 517 LKRTIAELPLTIAPEVFTSTYPMIGLRTGATHTLTLYTLVANVSTSGYLISCAS 576
OY 554 PTFHMASFFSNALYNSFYLAGFMINLSLMTVPAMISKVSELR---CFEGIMAKTQFS- 609
Db 577 SSISMALSVGPVPIYPIELFEGFLNSAS---VPYFYIYLSLMSFRYANALLIMDMST 633
OY 610 -----RRTYKMPGLNLTIAVSGDKILSAMELDYPYAIYLYIGSGFMYLY 658
Db 634 VVDEIACTRANWTCPSREILLETENFRV-EDFALDIACLFR--LIVIFRLGALLCJA 688

```

RESULT 11

WHIT_CERCA

ID WHIT_CERCA STANDARD; PRT: 679 AA.

AC 017320;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE White protein.

GN

OS Ceratitis capitata (Mediterranean fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OX Tephritidae; Tephritidae; Ceratitis.

OX NCBI_Taxid=7213;

RN SEQUENCE FROM N.A.

RX MEDLINE=96123276; PubMed=8533095;

RA Zviobell L.J., Saccone G., Zacharopoulou A., Besansky N.J.,

RA Pavla G., Collins P.H., Louis C., Kafatos F.C.;

RT "The white gene of Ceratitis capitata: a phenotypic marker for

RT germline transformation."

RL Science 270:2005-2007(1995).

CC -1- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM

CC NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT

CC CELLS RESPONSIBLE FOR EYE COLOR.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to license@sib-sib.ch).

CC -----

CC EMBL: X89933; CA61998.1; -

CC InterPro: IPR003593; AAA_ATPase.

CC InterPro: IPR003439; ABC_transporter.

CC Pfam: PF00005; ABC_tran.1.

CC ProDom: PD000006; ABC_transporter.1.

CC SMART: SM00387; AAA.1.

CC TIGRfam: TIGR00955; Jao1204.1.

CC PROSITE: PS00211; ABC_TRANSPORTER_1; 1.

CC PROSITE: PS00893; ABC_TRANSPORTER_2; 1.

CC Pigment: ATP-binding; Transmembrane; Transport

CC NP_BIND 121 128 ATP (BY SIMILARITY).

CC FT TRANSMEM 427 445 POTENTIAL.

```

FT TRANSMEM 457 477 POTENTIAL.
FT TRANSMEM 507 525 POTENTIAL.
FT TRANSMEM 534 555 POTENTIAL.
FT TRANSMEM 568 586 POTENTIAL.
FT TRANSMEM 651 670 POTENTIAL.
FT CARBOHYD 628 628 POTENTIAL.
FT CARBOHYD 643 643 N-LINKED (GLCNAc. . .) (POTENTIAL).
SQ SEQUENCE 679 AA: 75145 MW: 3F9CBCT8A835C4CC CRC64;

```

Query Match 17.8%; Score 623.5; DB 1; Length 679;
 Best Local Similarity 28.3%; Pred. No. 3, 2e-39;
 Matches 169; Conservative 112; Mismatches 264; Indels 53; Gaps 8;

```

OY 88 IONSFKVRSQMLAIIGSSCGRASLDVITGRHGCG-KIKSGOI-WINQPSSPOLVR 145
Db 104 LKNDSGAVYPCPELLAVMSSGAGKTTLLNAAFNSKVOQSPSTIRLNHPDAAKMQ 163
OY 146 KCVAHVRHNDLRLTYRETLATLAQRDLPTTSQADRVEDVLAELRQCADTRV 205
Db 164 ARCAVQDDLEFISLAREHLFOAMVRMFRHMTOKOKVQRVDVODLSLGCONTLI 223
OY 206 G-NMYVRGLSGERRRVSIGQLMNPGLILDEPTSGDSFTAHNLVKTLRLAKGNRL 264
Db 224 GVGPRVYGLSGEKKRLAFASALTDPELLDCDEPTSGDSFMAHVSVOYKLKLGKGT 283
OY 265 VLSLHOPRSDIFPLDVLVLTMTSGTPYLCAQHMVOYFALIGPCPRVSNPADFTYDL 324
Db 284 VILTIHQPSSLEFLFKIILMAGRAFLGTPEAVDFSYGATCPNTYTPADFTYVOV 343
OY 325 TS-----IDRSDEQL-----ATREKAOSLALFLEK--VRDLDDFLKAKETKD 367
Db 344 LAVVPGREVSRRVAKICDNFAVGKVSREMEQNFQVLVNSNGCKDEMEYTKASW-- 401
OY 368 LDEDTVESSVTPLDTNCLSPPTKMPAVQOFTLLIRQISNDPRDLPTLLIHGAECIM 427
Db 402 -----FMQFRAVLMRSMVLYLKEPLAVKRLQTTWV 433
OY 428 SMTIGLYFGHGSIOLEFMDTALLFMIGALIPFNVLIDVTSKCYSERAMLYELEDGLY 487
Db 434 AVLIGLIFLGOQLTOYGVNMINGAIFLFLNMTFQNSFATITFTLPEVMEETRSLRX 493
OY 488 TTPPYTPAKLGLPEHCAYIIIGMPYWMANLRPGLOPPLFLVLMVLFECRRIMAL 547
Db 494 RCDTYFLGKTALPLPLVPEFLETAIVPLILRGVDHFFALALVTLVANVSTSGF 553
OY 548 AAALPLTFHMASFFSNALYNSFYLAGFMINLSLMTVPAMISKVSELR---CFEGIMAKTQ 607
Db 554 IISCASSSTSMALSVGPVPIYPIELFEGFLNSGVPYVFKMLSYLSMFRYANALLIMDMST 633
OY 608 FS-----RRTYKMPGLNLTIAVSGDKILSAMELDYPYAIYLYIGSGFMYLYYSL 662
Db 614 MADVKRGEITCTLSNTCPSSGEVILETINFASDLPPFGLALLIVGFRISAYIAL 671

```

RESULT 12

ABG1_MOUSE

ID ABG1_MOUSE STANDARD; PRT: 666 AA.

AC 064343;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE ATP-binding cassette, sub-family G, member 1 (White protein homolog)

GN ABCG1 OR ABC8 OR WHRT.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_Taxid=10090;

RN SEQUENCE FROM N.A.

RX MEDLINE=97186700; PubMed=9034316;

RA Croop J.M., Tiller G.E., Fletcher J.A., Lux M.L., Raab E.,

RA Goldenson D., Son D., Arciniegas S., Wu R.;

RT *Isolation and characterization of a mammalian homolog of the
 RT Drosophila white gene.*;
 RL Gene 185:77-85(1997).
 RN (2)
 RC SEQUENCE FROM N.A.
 RA STRAIN-DBA/2;
 RX MEDLINE-96359154; PubMed-8703120;
 RA Savary S., Denizot F., Luciani M.-F., Mattei M.-G., Chimiini G.;
 RT *Molecular cloning of a mammalian ABC transporter homologous to
 RT Drosophila white gene.*;
 RL Mamm. Genome 7:673-676(1996).
 RN (3)
 RP SEQUENCE FROM N.A.
 RX MEDLINE-21092576; PubMed-11162488;
 RA Lorkowski S., Rust S., Engel T., Jung E., Tegelkamp K., Galinski E.A.,
 RA Assmann G., Cullen P.;
 RT *Genomic sequence and structure of the human ABCG1 (ABCG8) gene.*;
 RL Biochem. Biophys. Res. Commun. 280:121-131(2001).
 RN (4)
 RP INDUCTION, AND PROBABLE FUNCTION.
 RX MEDLINE-20261604; PubMed-1079558;
 RA Venkateswaran A., Repa J.J., Lobbaccaro J.-M.A., Bronson A.,
 RA Mangelsdorf D.J., Edwards P.A.;
 RT *Human white/murine ABC8 mRNA levels are highly induced in
 RT lipid-loaded macrophages. A transcriptional role for specific
 RT oxysterols.*;
 RL J. Biol. Chem. 275:14700-14707(2000).
 RN (5)
 RP REVIEW.
 RX MEDLINE-21474438; PubMed-11590207;
 RA Schmitz G., Langmann T., Heimerl S.;
 RT *Role of ABCG1 and other ABCG family members in lipid metabolism.*;
 RL J. Lipid Res. 42:1513-1520(2001).
 CC -1- FUNCTION: Transporter involved in macrophage lipid homeostasis. Is
 CC also be involved in intracellular lipid export complex. Could
 CC role in cellular lipid homeostasis may not be limited to
 CC macrophages.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- TISSUE SPECIFICITY: EXPRESSED MAINLY IN BRAIN, THYMUS, LUNG,
 CC ADRENALS, SPLEEN AND PLACENTA. LITTLE OR NO EXPRESSION IN LIVER,
 CC KIDNEY, HEART, MUSCLE OR TESTES.
 CC -1- INDUCTION: Strongly induced in macrophage cell line RAW264.7
 CC during cholesterol influx. Induction is mediated by the liver X
 CC receptor/cellnoid X receptor (LXR/RXR) pathway.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 CC SUPERFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: U34920; AAB47738.1; -
 CC DR EMBL: 248745; CA88636.1; -
 CC DR EMBL: AF323659; AAK27442.1; -
 CC DR MGD: MG1:107704; Abcgl.
 CC DR InterPro: IPR003593; AAA_Atpase.
 CC DR InterPro: IPR003439; AAA_Transporter.
 CC DR InterPro: IPR005284; P1gment_permease.
 CC Pfam: PF00005; ABC_tran.1.
 CC ProDom: PD000006; ABC_transporter.1.
 CC DR SMART: SM00382; AAA.1.
 CC DR TIGRfam: TIGR00955; 3a01204; 1.
 CC DR PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
 CC DR PROSITE: PS00893; ABC_TRANSPORTER_2; 1.
 CC KY ATP-binding; Transmembrane; Transport.
 CC FT DOMAIN 1 414 CYTOPLASMIC (POTENTIAL).
 CC TRANSMEM 415 433 POTENTIAL.
 CC FT DOMAIN 434 444 EXTRACELLULAR (POTENTIAL).
 CC

FT TRANSMEM 445 465 POTENTIAL.
 FT DOMAIN 466 494 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 495 513 POTENTIAL.
 FT DOMAIN 514 521 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 522 543 POTENTIAL.
 FT DOMAIN 544 555 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 556 574 POTENTIAL.
 FT DOMAIN 575 637 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 638 657 POTENTIAL.
 FT DOMAIN 658 666 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 118 125 ATP (POTENTIAL).
 SQ SEQUENCE 666 AA; 74033 MW; EDDCAFEFD4395086 CRC64;
 Query Match 17.7%; Score 621; DB 1; Length 666;
 Best Local Similarity 25.4%; Pred. No. 4.8e-39;
 Matches 165; Conservative 137; Mismatches 279; Indels 68; Gaps 16;
 QY 33 DNSLYFT---YSGQPN---TLEVDLNTVOVLASQVPEFEOIAQFKMPTSPSCONSC 84
 DB 57 DNN--FTEAQRFSSLPRAAVNIEFQDLSTV---DEGPWKKKRYKTL----- 100
 QY 85 ELGIQNLSEKVRSGOMLAIIGSSGCRASLLDVTGKGKIKSGQIWTNGOPSSPOLY 144
 DB 101 ---LKGISGFNGSELVATMGPSGACKSTIMILAGYREG--MKGAVLNGMPRLDLCF 155
 QY 145 RKCVAHYRQHNOLLPLVRETLAFTAQMRPLPTFSQAOADKREVDVIAELRLQCADTR 204
 DB 156 RKVCYIMDDMLPLHVTQEAAMVSAHLKQ--KDEGRREVKELRLGLPCANFR 213
 QY 205 VGNMYVRGSLGGERRVSIGVOLLMPGILLIDEPISGDSFPAHNLVYTLSLANGNRL 264
 DB 214 TGS-----LSGGKRLALALELVNPPWFDEPTEGSDSCFQVSLMKGLAOGGRS 268
 QY 265 VLISLHQPSPDIFRLDLVLTMTSGPIYGAOHVQVFTAGYCPRYSNPADRYVDL 324
 DB 269 IVCTIHQPSAKLELFDOLYVLSOGQVYRGKSNLPLRLDLGLMCPYHNADRVMEY 328
 QY 325 TSDIRSRQDELATREKAOSLAALFLEKY---RDLDPLMKAETKDLDEDTVESSVTP 381
 DB 329 ASGEYGDONSRLVRAVRECMCADYKRDIGDTPVPLMH---RAEDSASMECSHF 385
 QY 382 DINCPLSPFKMGAVQVFTTLIRQISNDPRDLPTLLIGAECLMSMTIGLYRGHSI 441
 DB 386 SASCL-----TQFCLERKRFSLMSRQSVLHRTSHIGLIGLIGLYLGIGNE 435
 QY 442 QLSFMDTALLFMGALIFPNVLIIVISCSYBRALVYELDGLYTTPTPFPAKTLGL 501
 DB 436 AKKVLNSGFLFFSMLEFPAALMPVLTPLDMSVFLREHLNMYSLKAVYIAKTMADV 495
 QY 502 PEHCAYIIYGMPTVLANLRGLOPFLHPLVLMVYVCCRIMALAAALPTFEMASF 561
 DB 496 PQIMEPPVAYCSIVYMTQSPDAVREVLALGVTSLVAOSLIGLIGAATSLQVAF 555
 QY 562 FSNALYNSYLLAGGPRINISLMTVPA---WISKVSFLMWCCEGLMKTQF--SRRTYKP 616
 DB 556 VGPVTAIPVLLSFGFVSFD---TTPVQNMYSISVYVYGEGLISYIGDRDLHCD 612
 QY 617 LGNLTIAVSGDKILSAMELDYPVLAIVYIGLSGGFVNLVYSLRF 665
 DB 613 IAECHFQKSEALRELDVENAKLY-LDFEVLG-----IFPISLRIL 653
 RESULT 13
 WHIT_LUCCU
 ID WHIT_LUCCU STANDARD; PRT; 677 AA.
 AC 005360;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE White protein.
 OS Lucilia cuprina (Greenbottle fly) (Australian sheep blowfly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera: Endopterygota: Diptera; Brachycera: Muscomorpha: Oestroidea;
 OC Calliphoridae: Lucilla.
 OX NCBI_Taxid=7375;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97087158; PubMed=8933176;
 RA Garcia R.L., Perkins H.D., Howells A.J.:
 RT "The structure, sequence and developmental pattern of expression of
 the white gene in the blowfly *Lucilla cuprina*.";
 RL Insect Mol. Biol. 5:251-260(1996).
 RN (2)
 RP SEQUENCE OF 490-584 FROM N.A.
 RX MEDLINE=90264941; PubMed=1971656;
 RA Elieur A., Vacek A.T., Howells A.J.:
 RT "Cloning and characterization of the white and topaz eye color genes
 from the sheep blowfly *Lucilla cuprina*.";
 RL J. Mol. Evol. 30:347-358(1990).
 CC -1- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
 NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
 CELLS RESPONSIBLE FOR EYE COLOR.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC This SMSS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U38899; AAA82057.1; -
 DR EMBL: X53265; CA337365.1; -
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transporter.
 DR InterPro: IPR005284; Pigment_permease.
 DR Pfam: PF00005; ABC_tran.1
 DR ProDom: PD000006; ABC_transporter.1.
 DR SMART; SM00382; AAA.1.
 DR TIGRfams: TIGR00955; 3a01204; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE: PS00893; ABC_TRANSPORTER_2; 1.
 RM Pigment: ATP-binding; Transmembrane; Transport.
 FT NP_BIND 119 126
 FT TRANSMEM 431 451 POTENTIAL.
 FT TRANSMEM 456 476 POTENTIAL.
 FT TRANSMEM 506 526 POTENTIAL.
 FT TRANSMEM 534 554 POTENTIAL.
 FT TRANSMEM 563 583 POTENTIAL.
 FT TRANSMEM 647 667 POTENTIAL.
 SQ SEQUENCE 677 AA; 75365 MW; D16FC11C97ED51D CRC64;
 Query Match 17.7%; Score 620.5; DB 1; Length 677;
 Best Local Similarity 29.3%; Pred. NO. 5.4e-39;
 Matches 174; Conservative 119; Mismatches 259; Indels 41; Gaps 13;
 OY 88 IONLSFKRSGOMLAITSSCCGRASLDVYTCR-GHCCKTKSGOI-WINQOPSSPOLVR 145
 DB 102 IKMCGVAVPELLAVMGSSGAGKTTLLNALAFSARGVOI-SPSSVRMLNGHPYDAKEMQ 161
 OY 146 KCAVHRQNLPLPLVRETIAFAOMRLPRTSQORMDRVEVYIELRLQCAQDVRV 205
 DB 162 ARCAVVOODLFIGSLTAREHLITQAVTMRPTMKOKLQKLRVDQVTDLSLKQNTI 221
 OY 206 G-NMIVRGLSGGERRRVSIGVQLMNPGLILDEPTSGLDSTFTHNLVTKTSLRKLKGRNL 264
 DB 222 GVPCKVKGLSGGERKRLAFASALDPLLCIDDEPTSGLDSTFMAASVQVYKLSQKRT 281
 OY 265 VLSLHQRSDIFRLFDVLVMTSGTPTIYLCAQAHMOYFPAIGPCPRISNPADFYVDL 324
 DB 282 VILTHQSSSELPFELDKLLMAGRAVAFGLTPEAVDVFESFISGOCPTNNPADFYQV 341
 OY 325 TSIDRRSREGELATREKAQSLALF-LEKV-RDLDLFLMK--AEFKDDEPTCVSSSTVP 380

DB 342 LAV---VPGREIESDRISKIDCNFAYGVSRHEMOQKIAKTDGLQKOD----- 390
 OY 381 LPTNCLPSPTKMPGAVOQFTTLIRROISNDFDPLTLIHGAEALMSTGTFLXFGHS 440
 DB 391 -ETTLIKRSMF-----TOFRAIMMSWISTLEPLLVKRLIQTTWVAVALGLILNDPM 445
 OY 441 IOLSEMDPAALLFMIGALIPFVILDVISKCYSEBAMLYELEDGLYTTGPFPAKILGE 500
 DB 446 TQGVWNNINGALFLELTLNMTFQNVAVINVFSELPVFKMRSLRKYRDTYFLGKTLAE 505
 OY 501 LPEHCAYIIYGMPTWLANLRPGLOPFLHFLVWLWVFCRRIMALAAALPTFHMS 560
 DB 506 LPLFLVLPFLFAIYAVPMGLRPGITHEFLSALALVTLVANVSTSGYLISCASTSTSMAL 565
 OY 561 FESNALYNFYLAGEFMINLSLWTPVAMISVSLRCMCFBEMLIQSRRTYKMPLC-- 618
 DB 566 SVGPPLTLPFLLEGLFVGVFNVGSVPVYFKMLSTFSWFRANEGLLINQMA---DVQPGEI 621
 OY 619 -----NLAVSGDKILSAMELD---SYPLVYIVLIVIGLGGFVWLYVSLR 663
 DB 622 TCTSTNTCPSSGXVXLETINFRDKFTRLIGLILIL-----IFRIAGTVAAK 670
 RESULT 14
 ID ABG1_HUMAN STANDARD; PRT; 678 AA.
 AC P45844; Q9BXK6; Q9BXK7; Q9BXK8; Q9BXK9; Q9BXL0; Q9BXL1; Q9BXL2;
 AC Q9BXL3; Q9BXL4;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 1 (white protein homolog)
 DE (ATP-binding cassette transporter 8).
 GN ABCG1 OR ABCG8 OR WHR1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN (1)
 RP SEQUENCE OF 3-678 FROM N.A. (ISOFORMS 1 AND 4).
 RC TISSUE:Retina.
 RX MEDLINE=96256850; PubMed=8659545;
 RA Chen H.M., Rossier C., Laliot M.D., Lynn A., Chakravarti A.,
 RA Perrin G., Antonarakis S.E.;
 RT "Cloning of the cDNA for a human homologue of the *Drosophila* white
 RL gene and mapping to chromosome 21q22.3.";
 RL Am. J. Hum. Genet. 59:66-75(1996).
 RN (2)
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
 RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
 RA Polley A., Menzel U., Delabar J., Kump K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Matsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordsiek G., Horstschler K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
 RA Ramser J., Beck A., Klages S., Hennig S., Rlesseimann L., Dagand E.,
 RA Weimayer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Leinhardt H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 RN (3)
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=20408883; PubMed=10950923;
 RA Berry A., Scott H.S., Kudoh J., Tallor I., Korostishevsky M.,
 RA Wattenhofer M., Gulpioni M., Barras C., Rossier C., Shibusky M.,
 RA Wang J., Kawasaki K., Asakawa S., Minoshima S., Shimizu N.,
 RA Antonarakis S.E., Bonne-Tamir B.;
 RT "Refined localization of autosomal recessive nonsyndromic deafness

RT DPM10 locus using 34 novel microsatellite markers, genomic
RT structure, and exclusion of six known genes in the region."
RL Genomics 68:22-29(2000).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE-21192304; PubMed-11279031;
RA Porsch-Oezcuenermez M., Langmann T., Helmerl S., Borsukova H.,
RA Kaminski W.E., Drobnik W., Honer C., Schumacher C., Schmitz G.;
RT "The zinc finger protein 202 (ZNF202) is a transcriptional repressor
RT of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene
RT expression and a modulator of cellular lipid efflux."
RL J Biol. Chem. 276:12427-12433(2001).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4; 5; 6 AND 7).
RX MEDLINE-21092576; PubMed-11162488;
RA Lorkowski S., Rust S., Engel T., Jung E., Tegelkamp K., Galinski E.A.,
RA Assmann G., Cullen P.;
RT "Genomic sequence and structure of the human ABCG1 (ABCG8) gene."
RL Biochem. Biophys. Res. Commun. 280:121-131(2001).
RN [6]
RP SEQUENCE OF 33-678 FROM N.A.
RX TISSUE-Petal brain;
RC MEDLINE-97166700; PubMed-9034316;
RA Croop J.M., Tiller G.E., Fletcher J.A., Lux M.L., Raab E.,
RA Goldenson D., Archilegas S., Son D., Wu R.;
RT "Isolation and characterization of a mammalian homolog of the
RT Drosophila white gene."
RL Gene 185:77-85(1997).
RN [7]
RP INDUCTION, AND PROBABLE FUNCTION.
RX MEDLINE-20261604; PubMed-10799558;
RA Venkateswaran A., Repa J.J., Lobbaccaro J.-M.A., Bronson A.,
RA Mangelsdorf D.J., Edwards P.A.;
RT "Human white/murine ABC8 mRNA levels are highly induced in
RT lipid-loaded macrophages. A transcriptional role for specific
RT oxysterols."
RL J. Biol. Chem. 275:14700-14707(2000).
RN [8]
RP INDUCTION, AND PROBABLE FUNCTION.
RX MEDLINE-20105556; PubMed-10639163;
RA Klucken J., Buechler C., Orso E., Kaminski W.E.,
RA Porsch-Oezcuenermez M., Liebisch G., Kapinsky M., Diederich W.,
RA Drobnik W., Dean M., Allikmets R., Schmitz G.;
RT "ABCG1 (ABCG8), the human homolog of the Drosophila white gene, is a
RT regulator of macrophage cholesterol and phospholipid transport."
RL Proc. Natl. Acad. Sci. U.S.A. 97:817-822(2000).
RN [9]
RP REVIEW.
RX MEDLINE-21474438; PubMed-11590207;
RA Schmitz G., Langmann T., Helmerl S.;
RT "Role of ABCG1 and other ABCG family members in lipid metabolism."
RL J. Lipid Res. 42:1513-1520(2001).
RN -1-
RP FUNCTION: Transporter involved in macrophage lipid homeostasis. Is
CC also be involved in intracellular lipid export complex. Could
CC role in cellular lipid homeostasis may not be limited to
CC macrophages.
CC -1-
RP SUBUNIT: May form heterodimers with several heterologous partners
CC of the ABCG subfamily.
CC -1-
RP SUBCELLULAR LOCATION: Integral membrane protein. Predominantly
CC localized in the intracellular compartments mainly associated with
CC the endoplasmic reticulum (ER) and Golgi membranes.
CC -1-
RP ALTERNATIVE PRODUCTS:
CC Event-alternative splicing. Named isoforms-7;
CC Comment-Additional isoforms seem to exist;
CC Name-1;
CC IsoId-P45844-1; Sequence-Displayed;
CC Name-2; Synonyms-4;
CC IsoId-P45844-2; Sequence-VSP_000047, VSP_000051;
CC Name-3; Synonyms-ABDE;
CC IsoId-P45844-3; Sequence-VSP_000048, VSP_000051;
CC Name-4; Synonyms-G;
CC IsoId-P45844-4; Sequence-VSP_000051;

CC Name-5; Synonyms-F;
CC IsoId-P45844-5; Sequence-VSP_000049, VSP_000051;
CC Name-6; Synonyms-HI;
CC IsoId-P45844-6; Sequence-VSP_000046, VSP_000051;
CC Name-7; Synonyms-C;
CC IsoId-P45844-7; Sequence-VSP_000050, VSP_000051;
CC -1-
RP TISSUE SPECIFICITY: EXPRESSED IN SEVERAL TISSUES.
CC -1-
RP INDUCTION: Strongly induced in monocyte-derived macrophages during
CC cholesterol influx. Conversely, mRNA and protein expression are
CC suppressed by lipid efflux. Induction is mediated by the liver x
CC receptor/retinoid x receptor (LXR/RXR) pathway.
CC -1-
RP SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X91249; CAA62631.1; ALT_INIT.
DR EMBL: AP001746; BAA95530.1; ALT_INIT.
DR EMBL: AB038161; BAB13728.2; ALT_INIT.
DR EMBL: AJ289137; CAC00730.1; JOINED.
DR EMBL: AJ289138; CAC00730.1; JOINED.
DR EMBL: AJ289139; CAC00730.1; JOINED.
DR EMBL: AJ289140; CAC00730.1; JOINED.
DR EMBL: AJ289141; CAC00730.1; JOINED.
DR EMBL: AJ289142; CAC00730.1; JOINED.
DR EMBL: AJ289143; CAC00730.1; JOINED.
DR EMBL: AJ289144; CAC00730.1; JOINED.
DR EMBL: AJ289145; CAC00730.1; JOINED.
DR EMBL: AJ289146; CAC00730.1; JOINED.
DR EMBL: AJ289147; CAC00730.1; JOINED.
DR EMBL: AJ289148; CAC00730.1; JOINED.
DR EMBL: AJ289149; CAC00730.1; JOINED.
DR EMBL: AJ289150; CAC00730.1; JOINED.
DR EMBL: AJ289151; CAC00730.1; JOINED.
DR EMBL: AF323658; AAK28836.1; JOINED.
DR EMBL: AF323659; AAK28836.1; JOINED.
DR EMBL: AF323660; AAK28836.1; JOINED.
DR EMBL: AF323661; AAK28836.1; JOINED.
DR EMBL: AF323662; AAK28836.1; JOINED.
DR EMBL: AF323663; AAK28836.1; JOINED.
DR EMBL: AF323664; AAK28836.1; JOINED.
DR EMBL: AF323665; AAK28836.1; JOINED.
DR EMBL: AF323666; AAK28836.1; JOINED.
DR EMBL: AF323667; AAK28836.1; JOINED.
DR EMBL: AF323668; AAK28836.1; JOINED.
DR EMBL: AF323669; AAK28836.1; JOINED.
DR EMBL: AF323670; AAK28836.1; JOINED.
DR EMBL: AF323671; AAK28836.1; JOINED.
DR EMBL: AF323672; AAK28836.1; JOINED.
DR EMBL: AF323673; AAK28836.1; JOINED.
DR EMBL: AF323674; AAK28836.1; JOINED.
DR EMBL: AF323675; AAK28836.1; JOINED.
DR EMBL: AF323676; AAK28836.1; JOINED.
DR EMBL: AF323677; AAK28836.1; JOINED.
DR EMBL: AF323678; AAK28836.1; JOINED.
DR EMBL: AF323679; AAK28836.1; JOINED.
DR EMBL: AF323680; AAK28836.1; JOINED.
DR EMBL: AF323681; AAK28836.1; JOINED.
DR EMBL: AF323682; AAK28836.1; JOINED.
DR EMBL: AF323683; AAK28836.1; JOINED.
DR EMBL: AF323684; AAK28836.1; JOINED.
DR EMBL: AF323685; AAK28836.1; JOINED.
DR EMBL: AF323686; AAK28836.1; JOINED.
DR EMBL: AF323687; AAK28836.1; JOINED.
DR EMBL: AF323688; AAK28836.1; JOINED.
DR EMBL: AF323689; AAK28836.1; JOINED.
DR EMBL: AF323690; AAK28836.1; JOINED.
DR EMBL: AF323691; AAK28836.1; JOINED.
DR EMBL: AF323692; AAK28836.1; JOINED.
DR EMBL: AF323693; AAK28836.1; JOINED.
DR EMBL: AF323694; AAK28836.1; JOINED.
DR EMBL: AF323695; AAK28836.1; JOINED.
DR EMBL: AF323696; AAK28836.1; JOINED.
DR EMBL: AF323697; AAK28836.1; JOINED.
DR EMBL: AF323698; AAK28836.1; JOINED.
DR EMBL: AF323699; AAK28836.1; JOINED.
DR EMBL: AF323700; AAK28836.1; JOINED.
DR EMBL: AF323701; AAK28836.1; JOINED.
DR EMBL: AF323702; AAK28836.1; JOINED.
DR EMBL: AF323703; AAK28836.1; JOINED.
DR EMBL: AF323704; AAK28836.1; JOINED.
DR EMBL: AF323705; AAK28836.1; JOINED.
DR EMBL: AF323706; AAK28836.1; JOINED.
DR EMBL: AF323707; AAK28836.1; JOINED.
DR EMBL: AF323708; AAK28836.1; JOINED.
DR EMBL: AF323709; AAK28836.1; JOINED.
DR EMBL: AF323710; AAK28836.1; JOINED.
DR EMBL: AF323711; AAK28836.1; JOINED.
DR EMBL: AF323712; AAK28836.1; JOINED.
DR EMBL: AF323713; AAK28836.1; JOINED.
DR EMBL: AF323714; AAK28836.1; JOINED.
DR EMBL: AF323715; AAK28836.1; JOINED.
DR EMBL: AF323716; AAK28836.1; JOINED.
DR EMBL: AF323717; AAK28836.1; JOINED.
DR EMBL: AF323718; AAK28836.1; JOINED.
DR EMBL: AF323719; AAK28836.1; JOINED.
DR EMBL: AF323720; AAK28836.1; JOINED.
DR EMBL: AF323721; AAK28836.1; JOINED.
DR EMBL: AF323722; AAK28836.1; JOINED.
DR EMBL: AF323723; AAK28836.1; JOINED.
DR EMBL: AF323724; AAK28836.1; JOINED.
DR EMBL: AF323725; AAK28836.1; JOINED.
DR EMBL: AF323726; AAK28836.1; JOINED.
DR EMBL: AF323727; AAK28836.1; JOINED.
DR EMBL: AF323728; AAK28836.1; JOINED.
DR EMBL: AF323729; AAK28836.1; JOINED.
DR EMBL: AF323730; AAK28836.1; JOINED.
DR EMBL: AF323731; AAK28836.1; JOINED.
DR EMBL: AF323732; AAK28836.1; JOINED.
DR EMBL: AF323733; AAK28836.1; JOINED.
DR EMBL: AF323734; AAK28836.1; JOINED.
DR EMBL: AF323735; AAK28836.1; JOINED.
DR EMBL: AF323736; AAK28836.1; JOINED.
DR EMBL: AF323737; AAK28836.1; JOINED.
DR EMBL: AF323738; AAK28836.1; JOINED.
DR EMBL: AF323739; AAK28836.1; JOINED.
DR EMBL: AF323740; AAK28836.1; JOINED.
DR EMBL: AF323741; AAK28836.1; JOINED.
DR EMBL: AF323742; AAK28836.1; JOINED.
DR EMBL: AF323743; AAK28836.1; JOINED.
DR EMBL: AF323744; AAK28836.1; JOINED.
DR EMBL: AF323745; AAK28836.1; JOINED.
DR EMBL: AF323746; AAK28836.1; JOINED.
DR EMBL: AF323747; AAK28836.1; JOINED.
DR EMBL: AF323748; AAK28836.1; JOINED.
DR EMBL: AF323749; AAK28836.1; JOINED.
DR EMBL: AF323750; AAK28836.1; JOINED.
DR EMBL: AF323751; AAK28836.1; JOINED.
DR EMBL: AF323752; AAK28836.1; JOINED.
DR EMBL: AF323753; AAK28836.1; JOINED.
DR EMBL: AF323754; AAK28836.1; JOINED.
DR EMBL: AF323755; AAK28836.1; JOINED.
DR EMBL: AF323756; AAK28836.1; JOINED.
DR EMBL: AF323757; AAK28836.1; JOINED.
DR EMBL: AF323758; AAK28836.1; JOINED.
DR EMBL: AF323759; AAK28836.1; JOINED.
DR EMBL: AF323760; AAK28836.1; JOINED.
DR EMBL: AF323761; AAK28836.1; JOINED.
DR EMBL: AF323762; AAK28836.1; JOINED.
DR EMBL: AF323763; AAK28836.1; JOINED.
DR EMBL: AF323764; AAK28836.1; JOINED.
DR EMBL: AF323765; AAK28836.1; JOINED.
DR EMBL: AF323766; AAK28836.1; JOINED.
DR EMBL: AF323767; AAK28836.1; JOINED.
DR EMBL: AF323768; AAK28836.1; JOINED.
DR EMBL: AF323769; AAK28836.1; JOINED.
DR EMBL: AF323770; AAK28836.1; JOINED.
DR EMBL: AF323771; AAK28836.1; JOINED.
DR EMBL: AF323772; AAK28836.1; JOINED.
DR EMBL: AF323773; AAK28836.1; JOINED.
DR EMBL: AF323774; AAK28836.1; JOINED.
DR EMBL: AF323775; AAK28836.1; JOINED.
DR EMBL: AF323776; AAK28836.1; JOINED.
DR EMBL: AF323777; AAK28836.1; JOINED.
DR EMBL: AF323778; AAK28836.1; JOINED.
DR EMBL: AF323779; AAK28836.1; JOINED.
DR EMBL: AF323780; AAK28836.1; JOINED.
DR EMBL: AF323781; AAK28836.1; JOINED.
DR EMBL: AF323782; AAK28836.1; JOINED.
DR EMBL: AF323783; AAK28836.1; JOINED.
DR EMBL: AF323784; AAK28836.1; JOINED.
DR EMBL: AF323785; AAK28836.1; JOINED.
DR EMBL: AF323786; AAK28836.1; JOINED.
DR EMBL: AF323787; AAK28836.1; JOINED.
DR EMBL: AF323788; AAK28836.1; JOINED.
DR EMBL: AF323789; AAK28836.1; JOINED.
DR EMBL: AF323790; AAK28836.1; JOINED.
DR EMBL: AF323791; AAK28836.1; JOINED.
DR EMBL: AF323792; AAK28836.1; JOINED.
DR EMBL: AF323793; AAK28836.1; JOINED.
DR EMBL: AF323794; AAK28836.1; JOINED.
DR EMBL: AF323795; AAK28836.1; JOINED.
DR EMBL: AF323796; AAK28836.1; JOINED.
DR EMBL: AF323797; AAK28836.1; JOINED.
DR EMBL: AF323798; AAK28836.1; JOINED.
DR EMBL: AF323799; AAK28836.1; JOINED.
DR EMBL: AF323800; AAK28836.1; JOINED.
DR EMBL: AF323801; AAK28836.1; JOINED.
DR EMBL: AF323802; AAK28836.1; JOINED.
DR EMBL: AF323803; AAK28836.1; JOINED.
DR EMBL: AF323804; AAK28836.1; JOINED.
DR EMBL: AF323805; AAK28836.1; JOINED.
DR EMBL: AF323806; AAK28836.1; JOINED.
DR EMBL: AF323807; AAK28836.1; JOINED.
DR EMBL: AF323808; AAK28836.1; JOINED.
DR EMBL: AF323809; AAK28836.1; JOINED.
DR EMBL: AF323810; AAK28836.1; JOINED.
DR EMBL: AF323811; AAK28836.1; JOINED.
DR EMBL: AF323812; AAK28836.1; JOINED.
DR EMBL: AF323813; AAK28836.1; JOINED.
DR EMBL: AF323814; AAK28836.1; JOINED.
DR EMBL: AF323815; AAK28836.1; JOINED.
DR EMBL: AF323816; AAK28836.1; JOINED.
DR EMBL: AF323817; AAK28836.1; JOINED.
DR EMBL: AF323818; AAK28836.1; JOINED.
DR EMBL: AF323819; AAK28836.1; JOINED.
DR EMBL: AF323820; AAK28836.1; JOINED.
DR EMBL: AF323821; AAK28836.1; JOINED.
DR EMBL: AF323822; AAK28836.1; JOINED.
DR EMBL: AF323823; AAK28836.1; JOINED.
DR EMBL: AF323824; AAK28836.1; JOINED.
DR EMBL: AF323825; AAK28836.1; JOINED.
DR EMBL: AF323826; AAK28836.1; JOINED.
DR EMBL: AF323827; AAK28836.1; JOINED.
DR EMBL: AF323828; AAK28836.1; JOINED.
DR EMBL: AF323829; AAK28836.1; JOINED.
DR EMBL: AF323830; AAK28836.1; JOINED.
DR EMBL: AF323831; AAK28836.1; JOINED.
DR EMBL: AF323832; AAK28836.1; JOINED.
DR EMBL: AF323833; AAK28836.1; JOINED.
DR EMBL: AF323834; AAK28836.1; JOINED.
DR EMBL: AF323835; AAK28836.1; JOINED.
DR EMBL: AF323836; AAK28836.1; JOINED.
DR EMBL: AF323837; AAK28836.1; JOINED.
DR EMBL: AF323838; AAK28836.1; JOINED.
DR EMBL: AF323839; AAK28836.1; JOINED.
DR EMBL: AF323840; AAK28836.1; JOINED.
DR EMBL: AF323841; AAK28836.1; JOINED.
DR EMBL: AF323842; AAK28836.1; JOINED.
DR EMBL: AF323843; AAK28836.1; JOINED.
DR EMBL: AF323844; AAK28836.1; JOINED.
DR EMBL: AF323845; AAK28836.1; JOINED.
DR EMBL: AF323846; AAK28836.1; JOINED.
DR EMBL: AF323847; AAK28836.1; JOINED.
DR EMBL: AF323848; AAK28836.1; JOINED.
DR EMBL: AF323849; AAK28836.1; JOINED.
DR EMBL: AF323850; AAK28836.1; JOINED.
DR EMBL: AF323851; AAK28836.1; JOINED.
DR EMBL: AF323852; AAK28836.1; JOINED.
DR EMBL: AF323853; AAK28836.1; JOINED.
DR EMBL: AF323854; AAK28836.1; JOINED.
DR EMBL: AF323855; AAK28836.1; JOINED.
DR EMBL: AF323856; AAK28836.1; JOINED.
DR EMBL: AF323857; AAK28836.1; JOINED.
DR EMBL: AF323858; AAK28836.1; JOINED.
DR EMBL: AF323859; AAK28836.1; JOINED.
DR EMBL: AF323860; AAK28836.1; JOINED.
DR EMBL: AF323861; AAK28836.1; JOINED.
DR EMBL: AF323862; AAK28836.1; JOINED.
DR EMBL: AF323863; AAK28836.1; JOINED.
DR EMBL: AF323864; AAK28836.1; JOINED.
DR EMBL: AF323865; AAK28836.1; JOINED.
DR EMBL: AF323866; AAK28836.1; JOINED.
DR EMBL: AF323867; AAK28836.1; JOINED.
DR EMBL: AF323868; AAK28836.1; JOINED.
DR EMBL: AF323869; AAK28836.1; JOINED.
DR EMBL: AF323870; AAK28836.1; JOINED.
DR EMBL: AF323871; AAK28836.1; JOINED.
DR EMBL: AF323872; AAK28836.1; JOINED.
DR EMBL: AF323873; AAK28836.1; JOINED.
DR EMBL: AF323874; AAK28836.1; JOINED.
DR EMBL: AF323875; AAK28836.1; JOINED.
DR EMBL: AF323876; AAK28836.1; JOINED.
DR EMBL: AF323877; AAK28836.1; JOINED.
DR EMBL: AF323878; AAK28836.1; JOINED.
DR EMBL: AF323879; AAK28836.1; JOINED.
DR EMBL: AF323880; AAK28836.1; JOINED.
DR EMBL: AF323881; AAK28836.1; JOINED.
DR EMBL: AF323882; AAK28836.1; JOINED.
DR EMBL: AF323883; AAK28836.1; JOINED.
DR EMBL: AF323884; AAK28836.1; JOINED.
DR EMBL: AF323885; AAK28836.1; JOINED.
DR EMBL: AF323886; AAK28836.1; JOINED.
DR EMBL: AF323887; AAK28836.1; JOINED.
DR EMBL: AF323888; AAK28836.1; JOINED.
DR EMBL: AF323889; AAK28836.1; JOINED.
DR EMBL: AF323890; AAK28836.1; JOINED.
DR EMBL: AF323891; AAK28836.1; JOINED.
DR EMBL: AF323892; AAK28836.1; JOINED.
DR EMBL: AF323893; AAK28836.1; JOINED.
DR EMBL: AF323894; AAK28836.1; JOINED.
DR EMBL: AF323895; AAK28836.1; JOINED.
DR EMBL: AF323896; AAK28836.1; JOINED.
DR EMBL: AF323897; AAK28836.1; JOINED.
DR EMBL: AF323898; AAK28836.1; JOINED.
DR EMBL: AF323899; AAK28836.1; JOINED.
DR EMBL: AF323900; AAK28836.1; JOINED.
DR EMBL: AF323901; AAK28836.1; JOINED.
DR EMBL: AF323902; AAK28836.1; JOINED.
DR EMBL: AF323903; AAK28836.1; JOINED.
DR EMBL: AF323904; AAK28836.1; JOINED.
DR EMBL: AF323905; AAK28836.1; JOINED.
DR EMBL: AF323906; AAK28836.1; JOINED.
DR EMBL: AF323907; AAK28836.1; JOINED.
DR EMBL: AF323908; AAK28836.1; JOINED.
DR EMBL: AF323909; AAK28836.1; JOINED.
DR EMBL: AF323910; AAK28836.1; JOINED.
DR EMBL: AF323911; AAK28836.1; JOINED.
DR EMBL: AF323912; AAK28836.1; JOINED.
DR EMBL: AF323913; AAK28836.1; JOINED.
DR EMBL: AF323914; AAK28836.1; JOINED.
DR EMBL: AF323915; AAK28836.1; JOINED.
DR EMBL: AF323916; AAK28836.1; JOINED.
DR EMBL: AF323917; AAK28836.1; JOINED.
DR EMBL: AF323918; AAK28836.1; JOINED.
DR EMBL: AF323919; AAK28836.1; JOINED.
DR EMBL: AF323920; AAK28836.1; JOINED.
DR EMBL: AF323921; AAK28836.1; JOINED.
DR EMBL: AF323922; AAK28836.1; JOINED.
DR EMBL: AF323923; AAK28836.1; JOINED.
DR EMBL: AF323924; AAK28836.1; JOINED.
DR EMBL: AF323925; AAK28836.1; JOINED.
DR EMBL: AF323926; AAK28836.1; JOINED.
DR EMBL: AF323927; AAK28836.1; JOINED.
DR EMBL: AF323928; AAK28836.1; JOINED.
DR EMBL: AF323929; AAK28836.1; JOINED.
DR EMBL: AF323930; AAK28836.1; JOINED.
DR EMBL: AF323931; AAK28836.1; JOINED.
DR EMBL: AF323932; AAK28836.1; JOINED.
DR EMBL: AF323933; AAK28836.1; JOINED.
DR EMBL: AF323934; AAK28836.1; JOINED.
DR EMBL: AF323935; AAK28836.1; JOINED.
DR EMBL: AF323936; AAK28836.1; JOINED.
DR EMBL: AF323937; AAK28836.1; JOINED.
DR EMBL: AF323938; AAK28836.1; JOINED.
DR EMBL: AF323939; AAK28836.1; JOINED.
DR EMBL: AF323940; AAK28836.1; JOINED.
DR EMBL: AF323941; AAK28836.1; JOINED.
DR EMBL: AF323942; AAK28836.1; JOINED.
DR EMBL: AF323943; AAK28836.1; JOINED.
DR EMBL: AF323944; AAK28836.1; JOINED.
DR EMBL: AF323945; AAK28836.1; JOINED.
DR EMBL: AF323946; AAK28836.1; JOINED.
DR EMBL: AF323947; AAK28836.1; JOINED.
DR EMBL: AF323948; AAK28836.1; JOINED.
DR EMBL: AF323949; AAK28836.1; JOINED.
DR EMBL: AF323950; AAK28836.1; JOINED.
DR EMBL: AF323951; AAK28836.1; JOINED.
DR EMBL: AF323952; AAK28836.1; JOINED.
DR EMBL: AF323953; AAK28836.1; JOINED.
DR EMBL: AF323954; AAK28836.1; JOINED.
DR EMBL: AF323955; AAK28836.1; JOINED.
DR EMBL: AF323956; AAK28836.1; JOINED.
DR EMBL: AF323957; AAK28836.1; JOINED.
DR EMBL: AF323958; AAK28836.1; JOINED.
DR EMBL: AF323959; AAK28836.1; JOINED.
DR EMBL: AF323960; AAK28836.1; JOINED.
DR EMBL: AF323961; AAK28836.1; JOINED.
DR EMBL: AF323962; AAK28836.1; JOINED.
DR EMBL: AF323963; AAK28836.1; JOINED.
DR EMBL: AF323964; AAK28836.1; JOINED.
DR EMBL: AF323965; AAK28836.1; JOINED.
DR EMBL: AF323966; AAK28836.1; JOINED.
DR EMBL: AF323967; AAK28836.1; JOINED.
DR EMBL: AF323968; AAK28836.1; JOINED.
DR EMBL: AF323969; AAK28836.1; JOINED.
DR EMBL: AF323970; AAK28836.1; JOINED.
DR EMBL: AF323971; AAK28836.1; JOINED.
DR EMBL: AF323972; AAK28836.1; JOINED.
DR EMBL: AF323973; AAK28836.1; JOINED.
DR EMBL: AF323974; AAK28836.1; JOINED.
DR EMBL: AF323975; AAK28836.1; JOINED.
DR EMBL: AF323976; AAK28836.1; JOINED.
DR EMBL: AF323977; AAK28836.1; JOINED.
DR EMBL: AF323978; AAK28836.1; JOINED.
DR EMBL: AF323979; AAK28836.1; JOINED.
DR EMBL: AF323980; AAK28836.1; JOINED.
DR EMBL: AF323981; AAK28836.1; JOINED.
DR EMBL: AF323982; AAK28836.1; JOINED.
DR EMBL: AF323983; AAK28836.1; JOINED.
DR EMBL: AF323984; AAK28836.1; JOINED.
DR EMBL: AF323985; AAK28836.1; JOINED.
DR EMBL: AF323986; AAK28836.1; JOINED.
DR EMBL: AF323987; AAK28836.1; JOINED.
DR EMBL: AF323988; AAK28836.1; JOINED.
DR EMBL: AF323989; AAK28836.1; JOINED.
DR EMBL: AF323990; AAK28836.1; JOINED.
DR EMBL: AF323991; AAK28836.1; JOINED.
DR EMBL: AF323992; AAK28836.1; JOINED.
DR EMBL: AF323993; AAK28836.1; JOINED.
DR EMBL: AF323994; AAK28836.1; JOINED.
DR EMBL: AF323995; AAK28836.1; JOINED.
DR EMBL: AF323996; AAK28836.1; JOINED.
DR EMBL: AF323997; AAK28836.1; JOINED.
DR EMBL: AF323998; AAK28836.1; JOINED.
DR EMBL: AF323999; AAK28836.1; JOINED.
DR EMBL: AF324000; AAK28836.1; JOINED.
DR EMBL: AF324001; AAK28836.1; JOINED.
DR EMBL: AF324002; AAK28836.1; JOINED.
DR EMBL: AF324003; AAK28836.1; JOINED.
DR EMBL: AF324004; AAK28836.1; JOINED.
DR EMBL: AF324005; AAK28836.1; JOINED.
DR EMBL: AF324006; AAK28836.1; JOINED.
DR EMBL: AF324007; AAK28836.1; JOINED.
DR EMBL: AF324008; AAK28836.1; JOINED.
DR EMBL: AF324009; AAK28836.1; JOINED.
DR EMBL: AF324010; AAK28836.1; JOINED.
DR EMBL: AF324011; A

DR EMBL; AF323658; AAK28835.1; JOINED.
 DR EMBL; AF323642; AAK28835.1; JOINED.
 DR EMBL; AF323645; AAK28835.1; JOINED.
 DR EMBL; AF323646; AAK28835.1; JOINED.
 DR EMBL; AF323647; AAK28835.1; JOINED.
 DR EMBL; AF323648; AAK28835.1; JOINED.
 DR EMBL; AF323649; AAK28835.1; JOINED.

Query Match 17.68; Score 617; DB 1; Length 678;
 Best Local Similarity 25.78; Pred. No. 9.8e-39;

Matches 173; Conservative 130; Mismatches 266; Indels 104; Gaps 18;

33 DNSLYFT--YSGOPN-----TLEVRDNLQVNDLASQVMPFQALQAFMPTSPSCQSCSEL 86
 57 DNNLEAOFSSLPRAAVNIFRDLSTV-----PESPMWRKKGYKTL----- 100.
 87 GIONLSFKVRSQOMLAIIGSSGCRASILDVITGRHGKIKSGQIWMINGOPSSPOLYVR 146
 101 -LKGISGRNSELVAIMPSCGASTLNNIAGYRGTG--KGANVLINGLRPDLRCFGR 157
 147 CVAHVRQHNQLPNTVRETTAFIAOMRLPRTFSQAQRDKRYEDVIAELRLQCADTRVG 206
 158 VSCYIMODMLPHTVDEAMVMSAHLKIQE--KDEGRREMYKEILLTALGLLSCANTRTG 215
 207 NMVYVGLSGGERRRVSIGVQLMNPGLILDEPTSGDSTFANHLVKTLSLAKGNRLV 266
 216 S-----LSGGQRRKRLAILELVNPPVDFEPTSGDASCFQVYSLMKGAQGRSII 270
 267 ISLHQRSDIFRLFDVLYLMTSGTFYLGAQOHVQYFAIYPCPRYSNPADFEYDLNS 326
 271 CIHQPSAKLFELFQLYLVLSGGQCYVRKVCNLPYLDLCLNCTYINPADFEYEVAS 330
 327 IDRRSRDEQL--ATEBKASIALALEKVRDI-----DQILMKAET-----KDL 369
 331 GEGYDQNSRLVAVRE-----GMCDSDHRRDLGGDAEYVNPFLMHRPSEVQKTRUKGLR 385
 370 EDTCYESSYTPDTCCLPSPFKMGVNOOFTLLIRROISNDPRDLPTLLIHGAECIMSM 429
 386 KDSSEMEGCHSFSACL-----TQFCILFKRTLSIMDSDVHLRLRTSHIGTL 435
 430 TIGFLYFGHSIOLSFMDTALLF-----MIGALIPFNVIIDVISKYSERAMLYELE 483
 436 LIGLILYIGINAKKVLNSGFLPFSMLFLMFAALP-----TVLFLPE 480
 484 DGL-----YTTGPFPAKILGELPEHCAYIIITGMPYTWLANLRGLOPFLIHFL 534
 481 MGVFLREHLNWTYSLKAYYLAKTMADVPOQIMEPAYCSIVYMTSOPSDAVRFVLEAL 540
 535 VMLVYFCRIMALAAALLPFFHMASFSSNALYNSEYLAGFMINLSLMTVPAMISKVS 594
 541 GTMISLVASIGLILGAASSTLOAVTFEYGPVTAIFVLLSFGFVSFDTPYTLQMWSTYS 600
 595 FLNCFEGIMKIQF--SRRTYKMPGLNLTIANVSGDKILSAMELDSYPLVAILIYIGLSG 652
 601 YVRVYFEGEVLISYIGLREDLHCDIDETCHPOKSAIILRELDVENAKLY-LDFIYLG--- 656
 653 GPMVLYYYSLEPI 665
 657 -----IFETSLRLI 665

RESULT 15
 YPC3_CAEEL
 ID YPC3_CAEEL STANDARD: PRT: 598 AA.
 AC Q11180;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Putative ABC transporter C05D10.3 in chromosome III.
 GN C05D10.3
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxId=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RA Du Z.
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP REVISIONS.
 RA Waterston R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; U13645; AAA20989.2;
 DR Wormpep: C05D10.3; CR29170.
 DR InterPro: IPR003593; AAA_Atpase.
 DR InterPro: IPR003439; ABC_transporter.
 DR InterPro: IPR005284; Pigment_permease.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transporter; 1.
 DR SMART: SM00382; AAA; 1.
 DR TIGRfam: TIGR00955; 3a01204; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER_1; FALSE NEG.
 DR PROSITE: PS00893; ABC_TRANSPORTER_2; 1.
 KW Hypothetical protein; ATP-binding; Transmembrane; Transport.
 FT NP_BIND 27 34 ATP (POTENTIAL).
 FT TRANSMEM 336 356 POTENTIAL.
 FT TRANSMEM 425 445 POTENTIAL.
 FT TRANSMEM 453 473 POTENTIAL.
 FT TRANSMEM 478 498 POTENTIAL.
 SQ SEQUENCE 598 AA: 66906 MW: 9D6414E06898E343 CRC64;

Query Match 17.18; Score 600; DB 1; Length 598;
 Best Local Similarity 27.98; Pred. No. 1.6e-37;
 Matches 170; Conservative 116; Mismatches 260; Indels 64; Gaps 15;

88 IONLSFKVRSQOMLAIIGSSGCRASILDVITGRHGKIKSGQIWMINGOPSSPOLYVR 147
 10 LHNVSAMESSGLLAIIGSSGKRTLNVLTRNDVGSSTLIDRRANKKIREM 69
 148 VAHVRQHNQLPNTVRETTAFIAOMRLPRTFSQAQRDKRYEDVIAELRLQCADTRVG 206
 149 I-----S----- 149
 70 SAFVOOHMFVGTAREHLQPMARLRMGDQYYSHERQLRVEQVLTQWGLKKCADTVIG 129
 207 -NMVYVGLSGGERRRVSIGVQLMNPGLILDEPTSGDSTFANHLVKTLSLAKGNRLV 265
 130 IPNQLKSGSGEKKRLSPASELTCTPKILCFDEPTSGDAPFAGVVALRSLAANGMTV 189
 266 LISTHQRSDIFRLFDVLYLMTSGTFYLGAQOHVQYFAIYPCPRYSNPADFEYDLT 325
 190 ITIHQPSHHVYSLFNNVCLMACGVIITLGRDQAVPFLPKCGYCPAYNPADHLIRL 249
 326 SIDRRSRDEQLATREKKAOSLAALFEKVRDIDFLMKAETDLDLDTG-----VES 376
 250 AVIDSDRATSMKT-----ISKIR--QGFV-----STDLGGSVALIGNANKLRAAS 292
 377 SVTPDTCCLPSPFKM-----PGAVOQFTLLIRROISNDPRDLPTLLIHGAECIMSM 429
 293 FVTGSDTS-----EKTTFPNODYNASEFTQFALFERSVLTVIRDPNLSVRLQTLITAF 349
 430 TIGFLYFGHSIOLSFMDTALLFMIIGALIPFNVIIDVISKYSERAMLYELE 481
 350 ITGIYEF-----QPTVPTATISINIM-FNHRNNKNEKLPNNVYITAEPIYARE 401
 482 LEDGLITGPFPAKILGELPEHCAYIIITGMPYTWLANLRGLOPFLIHFLVLYVFC 541

Db 402 NANGYRTSAIFLAKNIAELPOYIILPTILYNTIYWMGSLYPNEMNYCFASLVTLLITNV 461
OY 542 CRIMLAAALLPFFHMASFFSNALYNSFLAGGFMINLSSLMWYPAWISKVSPLRWCFE 601
Db 462 AISISYAVATIFANTDVAMTILPITFVVPIMAFGGFFETFDALPSYFKWLSSLSYFKGYE 521
OY 602 GLM-----KIQFSRRTYKMPGLNLTI--AVSGDKILSAMELD-SYPLXAYLIVIGLSG 652
Db 522 ALAINEMDSIKVIEPCFNSSWTAFALDSCPKNGHQVLESIDFSASHKTFDI-SILFGMTI 580
OY 653 GFMYLYYVSL 662
Db 581 GIRIIRAYVAL 590

Search completed: July 25, 2003, 17:12:45
Job time : 30 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein. - protein search, using sw model

Run on: July 25, 2003, 16:49:39 ; Search time 86 Seconds
(without alignments)
1242.128 Million cell updates/sec

Title: US-09-989-981a-8

Perfect score: 3506
Sequence: 1 MACKAAEERGLPKGATPPDT.....FVLYVSLFIRIKPSQDM 673

Scoring table: BLASTN62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3506	100.0	673	24	AAE31705
2	3502	99.9	673	23	ABP52129
3	2888.5	82.4	672	24	AAE31703
4	1961	55.9	374	23	ABG61539
5	730.5	20.8	632	21	AAE18079
6	730.5	20.8	648	21	AAE18078
7	724	20.7	625	21	AAE18080
8	713	20.3	652	23	AAU96986
9	705	20.1	651	23	AAU96990

10	697	19.9	651	23	AAU96984	Human ABCG5 proteol
11	697	19.9	651	23	AAU96993	Human ABCG5 proteol
12	697	19.9	651	23	AAE13290	Human ABCG5 proteol
13	697	19.9	651	24	AAE31704	Human ABCG5 proteol
14	696	19.9	651	23	AAU96989	Human ABCG5 proteol
15	694	19.8	651	23	AAU96992	Human ABCG5 proteol
16	691.5	19.7	652	23	AAU96985	Human ABCG5 proteol
17	688.5	19.6	652	23	AAE13289	Human ABCG5 proteol
18	688.5	19.6	652	23	AAE13308	Human ABCG5 proteol
19	688.5	19.6	652	23	AAE13309	Human ABCG5 proteol
20	688.5	19.6	652	24	AAE31702	Human ABCG5 proteol
21	675	19.3	649	23	ABP52128	Human ABCG5 proteol
22	666	19.0	657	23	ABP52127	Human ABCG5 proteol
23	666	19.0	657	23	ABP52128	Human ABCG5 proteol
24	656	18.7	657	21	AAE18081	Human ABCG5 proteol
25	642.5	18.3	655	22	AAE18082	Human ABCG5 proteol
26	642.5	18.3	655	22	AAE18083	Human ABCG5 proteol
27	642.5	18.3	655	23	AAE18084	Human ABCG5 proteol
28	640.5	18.3	655	23	AAE18085	Human ABCG5 proteol
29	640.5	18.3	655	23	AAE18086	Human ABCG5 proteol
30	638.5	18.2	655	21	AAE18087	Human ABCG5 proteol
31	638.5	18.2	655	21	AAE18088	Human ABCG5 proteol
32	638.5	18.2	655	22	AAE18089	Human ABCG5 proteol
33	638.5	18.2	655	23	AAE18090	Human ABCG5 proteol
34	638.5	18.2	655	23	AAE18091	Human ABCG5 proteol
35	638.5	18.2	655	23	AAE18092	Human ABCG5 proteol
36	634.5	18.1	655	23	AAE18093	Human ABCG5 proteol
37	621	17.7	666	23	AAE18094	Human ABCG5 proteol
38	620	17.7	666	23	AAE18095	Human ABCG5 proteol
39	618	17.6	666	23	AAE18096	Human ABCG5 proteol
40	617	17.6	666	23	AAE18097	Human ABCG5 proteol
41	612	17.5	674	23	AAE18098	Human ABCG5 proteol
42	610.5	17.4	674	23	AAE18099	Human ABCG5 proteol
43	572.5	16.3	609	22	AAE18100	Human ABCG5 proteol
44	571.5	16.3	646	24	AAE18101	Human ABCG5 proteol
45	570.5	16.3	646	24	AAE18102	Human ABCG5 proteol

ALIGNMENTS

RESULT 1	AAE31705	standard; Protein; 673 AA.
ID	AAE31705	
AC	AAE31705;	
DT	24-MAR-2003 (first entry)	
DE	Human ABCG8 proteol.	
XX		
XX		
KW	ABC family cholesterol transporter; ABCG8; sterol-related disorder;	
KW	silicosterolemia; hyperlipidemia; hypercholesterolemia; gall stone;	
KW	HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;	
KW	human; ATP-binding cassette; silicosterolemia susceptibility gene; SSG;	
XX		
XX		
OS	Homo sapiens.	
PN	WO200281691-A2.	
XX		
XX		
PD	17-OCT-2002.	
XX		
PF	20-NOV-2001; 2001WO-US43823.	
XX		
PR	20-NOV-2000; 2000US-252235P.	
XX		
XX	28-NOV-2000; 2000US-253645P.	
PA	(TULAR) TULARIK INC.	
XX	(TEXA) UNIV TEXAS SYSTEM.	
XX		
PI	Hobbs HH, Shan B, Barnes R, Tian H;	

DR WPI; 2003-058548/05.
 DR N-PSDB; AAD48883.
 XX
 PT New ABCG8 polypeptides and nucleic acids, useful for treating
 PT steroid-related disorders e.g. sitosterolemia, hypercholesterolemia,
 PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
 PT nutritional deficiencies -
 XX
 PS Claim 22; Page 81-82; 94pp; English.

XX
 CC The invention relates to ATP-binding cassette (ABC) family cholesterol
 CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
 CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also
 CC known as sitosterolemia susceptibility gene (SSG). Sequences of the
 CC invention are useful for treating or preventing sterol-related disorders
 CC such as sitosterolemia, hyperlipidemia, hypercholesterolemia, gall
 CC stones, HDL deficiency, atherosclerosis and nutritional deficiencies.
 CC They are also useful in gene therapy. The present sequence is human
 CC ABCG8 protein.
 XX
 SQ Sequence 673 AA:

Query Match 100.0%; Score 3506; DB 24; Length 673;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 673; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGKAAEERGLPKGATPOTSGLDRLFSSESNSLYFTYSGOPTLEVRDLNVOVDLAS 60
 DB 1 MAGKAAEERGLPKGATPOTSGLDRLFSSESNSLYFTYSGOPTLEVRDLNVOVDLAS 60
 QY 61 QVPWFEOALQFKMPTSPSCNSCELGIONLSFKVSGOMLAIIGSSCGRASLDVYTG 120
 DB 61 QVPWFEOALQFKMPTSPSCNSCELGIONLSFKVSGOMLAIIGSSCGRASLDVYTG 120
 QY 121 RGHGKIKSGQIWMINGPSSPOLYRKCAVHRQHNQLLPNTLVRETLAFTIOMRLPRTFS 180
 DB 121 RGHGKIKSGQIWMINGPSSPOLYRKCAVHRQHNQLLPNTLVRETLAFTIOMRLPRTFS 180
 QY 181 QAOORRVEDVIAELRLQCADTRVGNMYVRLSGGERRRVSIGVOLLMPGILILDEPT 240
 DB 181 QAOORRVEDVIAELRLQCADTRVGNMYVRLSGGERRRVSIGVOLLMPGILILDEPT 240
 QY 241 SGIDSTFAHNLVKTLSRLAKGNRLVLSIHOPRSIDFRFDVILMTSGTPYLGAAQHM 300
 DB 241 SGIDSTFAHNLVKTLSRLAKGNRLVLSIHOPRSIDFRFDVILMTSGTPYLGAAQHM 300
 QY 301 VOYFAIGYPCPRYSNPADFYVDLTSIDRSREOELATREKAQSLAALFEKVRDLDEL 360
 DB 301 VOYFAIGYPCPRYSNPADFYVDLTSIDRSREOELATREKAQSLAALFEKVRDLDEL 360
 QY 361 WKAETKRDLEDQCVSSVTPDLTNCPLPSTKAPGAVOQFTLIRROISNDFRDLPTLIH 420
 DB 361 WKAETKRDLEDQCVSSVTPDLTNCPLPSTKAPGAVOQFTLIRROISNDFRDLPTLIH 420
 QY 421 GAELACMSMTIGFLYFGHSIQLSFMDTAAALFMGALIPFVILDVISKCYSEKAMLY 480
 DB 421 GAELACMSMTIGFLYFGHSIQLSFMDTAAALFMGALIPFVILDVISKCYSEKAMLY 480
 QY 481 ELEDGLYTTGPFYFAKILGELPERHCAYIIIGMPTMYLANLRPGLOPPLFLVLYVF 540
 DB 481 ELEDGLYTTGPFYFAKILGELPERHCAYIIIGMPTMYLANLRPGLOPPLFLVLYVF 540
 QY 541 CCRIMAAALAAALLPFHNASFFSNALYNSFYLAGFMALNSLWTPVPMISKVSFLRMCF 600
 DB 541 CCRIMAAALAAALLPFHNASFFSNALYNSFYLAGFMALNSLWTPVPMISKVSFLRMCF 600
 QY 601 EGLMKIQFSRRTYKAPLGNLTIAVSGDKILSAMELDSTPLAIIILYIGLSGGFVLYYV 660
 DB 601 EGLMKIQFSRRTYKAPLGNLTIAVSGDKILSAMELDSTPLAIIILYIGLSGGFVLYYV 660
 QY 661 SLRPTKORPSQDW 673
 DB 661 SLRPTKORPSQDW 673

RESULT 2
 ID ABP52129 standard; Protein; 673 AA.
 XX
 AC ABP52129;
 XX

10-OCT-2002 (first entry)

Homo sapiens ABC transporter ABCG8 protein SEQ ID NO:81.

ATP-binding cassette transporter; ABC transporter; modulation; D loop;
 cancer; bacterial infection; fungal infection; protozoal infection;
 antibacterial; fungicide; protozoacide.

Homo sapiens.

EP1217066-A1.

26-JUN-2002.

21-DEC-2000; 2000EP-0870316.

21-DEC-2000; 2000EP-0870316.

(UYGE-) UNIV GENT.

WPI; 2002-550404/59.

Modulating activity of ATP-binding cassette (ABC) transporters by
 influencing dimerization of nucleotide binding domains through use of D
 loop sequence of an ABC transporter, or its antisense peptide or
 peptide mimetic -

Disclosure; Fig 3; 29pp; English.

The present invention describes a method (M1) for modulating the activity
 of ABC-binding cassette (ABC) transporters by influencing the
 dimerization of the nucleotide binding domains comprising using: (a) a
 polypeptide (polyp) consisting of 5-50 amino acids comprising the D loop
 sequence of an ABC transporter (ABP52049 to ABP52091); (b) a polyp
 consisting of the D loop sequence of (a) or (b); (c) a peptide
 mimetic or antisense peptide of (a) or (b). ABC transporters have
 antibacterial, fungicide and protozoacide activities. (M1) is useful for
 selectively modulating the activity of ABC transporters belonging to the
 group of multidrug transporter/P-glycoproteins. Bacterial, fungal or
 protozoal ABC transporters are involved in the infection of a mammal or
 in the induction of resistance to antibiotics or drugs in a mammal. (M1)
 is useful for preventing, treating or alleviating diseases associated
 with functionality of an ABC transporter. ABP52092 to ABP52140 represent
 ABC transporter proteins given in the exemplification of the present
 invention.

Sequence 673 AA:

Query Match 99.9%; Score 3502; DB 23; Length 673;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 672; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAGKAAEERGLPKGATPOTSGLDRLFSSESNSLYFTYSGOPTLEVRDLNVOVDLAS 60
 DB 1 MAGKAAEERGLPKGATPOTSGLDRLFSSESNSLYFTYSGOPTLEVRDLNVOVDLAS 60
 QY 61 QVPWFEOALQFKMPTSPSCNSCELGIONLSFKVSGOMLAIIGSSCGRASLDVYTG 120
 DB 61 QVPWFEOALQFKMPTSPSCNSCELGIONLSFKVSGOMLAIIGSSCGRASLDVYTG 120
 QY 121 RGHGKIKSGQIWMINGPSSPOLYRKCAVHRQHNQLLPNTLVRETLAFTIOMRLPRTFS 180
 DB 121 RGHGKIKSGQIWMINGPSSPOLYRKCAVHRQHNQLLPNTLVRETLAFTIOMRLPRTFS 180
 QY 181 QAOORRVEDVIAELRLQCADTRVGNMYVRLSGGERRRVSIGVOLLMPGILILDEPT 240


```

DB      |||||||
181  QAOQRKRVEDVIAELRLROCADTRVGNMYRGISGGERRRVSIQVOLLNPGILILDEPT 240
OY      241  SGLDSEFTANIKYKTSRLAKGNRLVLSIHQORSDFRLFDVLMTSGPTIYLGAAGHM 300
DB      241  SGLDSEFTANIKYKTSRLAKGNRLVLSIHQORSDFRLFDVLMTSGPTIYLGAAGHM 300
OY      301  VOYFTAIQPCPRYSNPADFYVDLTSIDRRSREOELATREKAOSLAALFLEKRDLDL 360
DB      301  VOYFTAIQPCPRYSNPADFYVDLTSIDRRSREOELATREKAOSLAALFLEKRDLDL 360
OY      361  WKAETKRDLEDTCEVSSVTPDLTNCPLSPFTKMGAVOQFTTLIRQISNDFRDLPTLLH 420
DB      361  WKAETKRDLEDTCEVSSVTPDLTNCPLSPFTKMGAVOQFTTLIRQISNDFRDLPTLLH 420
OY      421  GAELCLMSMTIGFLVFGHSIQLSFMDTAALFEMGALIPFNVIIDVISKYSERAMLY 480
DB      421  GAELCLMSMTIGFLVFGHSIQLSFMDTAALFEMGALIPFNVIIDVISKYSERAMLY 480
OY      481  ELEDGLYTTGPFYFAKIIIGELPEHCAYIIIGMPTWLANLRGLOPFLHLLVWLV 540
DB      481  ELEDGLYTTGPFYFAKIIIGELPEHCAYIIIGMPTWLANLRGLOPFLHLLVWLV 540
OY      541  CCRIMALAAALLPFFHNASFFSNALYNSFYLAGFMINLSSIMTVPAWISKVSFLRMCF 600
DB      541  CCRIMALAAALLPFFHNASFFSNALYNSFYLAGFMINLSSIMTVPAWISKVSFLRMCF 600
OY      601  EGLMTIOFSRRTYKMPGLMTAVSGDKILSMELDSYPLVAILVIGLSGGMVLYV 660
DB      601  EGLMTIOFSRRTYKMPGLMTAVSGDKILSMELDSYPLVAILVIGLSGGMVLYV 660
OY      661  SLRFKOKPSQDM 673
DB      661  SLRFKOKPSQDM 673

RESULT 3
AAE31703
ID      AAE31703 standard; Protein: 672 AA.
XX      AAE31703;
AC      AAE31703;
XX      24-MAR-2003 (first entry)
DE      Mouse ABCG8 protein.
XX      ABC family cholesterol transporter; ABCG8; sterol-related disorder;
KM      sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;
KM      HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;
KM      mouse; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;
KM      ABCG5.
OS      Mus sp.
XX      Key      Location/Qualifiers
XX      FT      Misc-difference 440
XX      /note= "Encoded by AAG"
XX      WO200281691-A2.
XX      17-OCT-2002.
XX      20-NOV-2001; 2001WO-US43823.
XX      20-NOV-2000; 2000US-252235P.
XX      28-NOV-2000; 2000US-253645P.
XX      PA      (TULAR) TULARIK INC.
XX      PA      (TEKA) UNIV TEXAS SYSTEM.
XX      Hobbs HH, Shan B, Barnes R, Tian H;
XX      WPI; 2003-058548/05.

```

```

DR      N-PSDB; A048881.
XX      PT      New ABCG8 polypeptides and nucleic acids, useful for treating
PT      sterol-related disorders e.g. sitosterolaemia, hypercholesterolaemia,
PT      hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
PT      nutritional deficiencies -
XX      PS      Claim 22; Page 76; 94pp; English.
XX      CC      The invention relates to ATP-binding cassette (ABC) family cholesterol
CC      transporter, ABCG8 polypeptides and polynucleotides. The invention also
CC      provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also
CC      known as sitosterolaemia susceptibility gene (SSG). Sequences of the
CC      invention are useful for treating or preventing sterol-related disorders
CC      such as sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall
CC      stones, HDL deficiency, atherosclerosis and nutritional deficiencies.
CC      They are also useful in gene therapy. The present sequence is mouse
CC      ABCG8 protein.
XX      SQ      Sequence      672 AA:
XX      Query Match      82.4%; Score 2888.5; DB 24; Length 672;
XX      Best Local Similarity 81.9%; Pred. No. 3.7e-287;
XX      Matches 551; Conservative 53; Mismatches 68; Indels 1; Gaps 1;
OY      1  MAGKAERGLPKGATPDTSGLQDRLESSESNSLYFTYSGOPTLEVRDINYOVDLAS 60
DB      1  MAEKTEKETQNLNNGVYLDASGLDLSFSESNSLYFTYSGQNTLEVRDLYQVDIAS 60
OY      61  QVWFEBQAOEFKIPMSHSSQDSCEGINSLSFKVRSQGMALISSGGGRSLDYVYG 120
DB      61  QVWFEBQAOEFKIPMSHSSQDSCEGINSLSFKVRSQGMALISSGGGRSLDYVYG 120
OY      121  RGHGKIKSGQIMINGOPSSPOLVRCVAVHQHQLNLTIVRETIATIAOMRLPRTFS 180
DB      121  RGHGKIKSGQIMINGOPSSPOLVRCVAVHQHQLNLTIVRETIATIAOMRLPRTFS 180
OY      181  QAOQRKRVEDVIAELRLROCADTRVGNMYRGISGGERRRVSIQVOLLNPGILILDEPT 240
DB      181  QAOQRKRVEDVIAELRLROCADTRVGNMYRGISGGERRRVSIQVOLLNPGILILDEPT 240
OY      241  SGLDSEFTANIKYKTSRLAKGNRLVLSIHQORSDFRLFDVLMTSGPTIYLGAAGHM 300
DB      241  SGLDSEFTANIKYKTSRLAKGNRLVLSIHQORSDFRLFDVLMTSGPTIYLGAAGHM 300
OY      301  VOYFTAIQPCPRYSNPADFYVDLTSIDRRSREOELATREKAOSLAALFLEKRDLDL 360
DB      301  VOYFTAIQPCPRYSNPADFYVDLTSIDRRSREOELATREKAOSLAALFLEKRDLDL 360
OY      361  WKAETKRDLEDTCEVSSVTPDLTNCPLSPFTKMGAVOQFTTLIRQISNDFRDLPTLLH 420
DB      361  WKAETKRDLEDTCEVSSVTPDLTNCPLSPFTKMGAVOQFTTLIRQISNDFRDLPTLLH 420
OY      421  GAELCLMSMTIGFLVFGHSIQLSFMDTAALFEMGALIPFNVIIDVISKYSERAMLY 480
DB      421  GAELCLMSMTIGFLVFGHSIQLSFMDTAALFEMGALIPFNVIIDVISKYSERAMLY 480
OY      481  ELEDGLYTTGPFYFAKIIIGELPEHCAYIIIGMPTWLANLRGLOPFLHLLVWLV 540
DB      481  ELEDGLYTTGPFYFAKIIIGELPEHCAYIIIGMPTWLANLRGLOPFLHLLVWLV 540
OY      541  CCRIMALAAALLPFFHNASFFSNALYNSFYLAGFMINLSSIMTVPAWISKVSFLRMCF 600
DB      541  CCRIMALAAALLPFFHNASFFSNALYNSFYLAGFMINLSSIMTVPAWISKVSFLRMCF 600
OY      601  EGLMTIOFSRRTYKMPGLMTAVSGDKILSMELDSYPLVAILVIGLSGGMVLYV 660
DB      601  EGLMTIOFSRRTYKMPGLMTAVSGDKILSMELDSYPLVAILVIGLSGGMVLYV 660
OY      661  SLRFKOKPSQDM 673
DB      661  SLRFKOKPSQDM 673

```

RESULT 4

ABG61539

ID ABG61539 standard; Protein: 374 AA.

AC ABG61539;

DT 27-AUG-2002 (first entry)

DE Human transporter and ion channel, TRICH9, Incyte ID 6585710CD1.

Human; transporter and ion channel; TRICH; transport disorder; neurological disorder; muscle disorder; immunological disorder; cancer; scleroderma; systemic lupus erythematosus; allergy; leukaemia; cell proliferative disorder; cervical cancer; breast cancer; neurodegenerative disorder; Parkinson's disease; Alzheimer's disease; myotonic dystrophy; catatonias; endocrine disorder; diabetes; Grave's disease; gastrointestinal disorder; Crohn's disease; renal disorder; Good pasture's syndrome; viral infection; cirrhosis; bacterial infection; fungal infection; parasitic infection; protozoal infection; helminthic infection; cardiovascular disorder; atherosclerosis; hepatic disease.

OS Homo sapiens.

PN MO200240541-A2.

PD 23-MAY-2002.

PF 25-OCT-2001; 2001MO-US46055.

PR 27-OCT-2000; 2000US-243989P.

PR 03-NOV-2000; 2000US-245904P.

PR 09-NOV-2000; 2000US-247673P.

PR 17-NOV-2000; 2000US-249661P.

PR 20-NOV-2000; 2000US-252232P.

PR 01-DEC-2000; 2000US-250790P.

XX (INCYTE GENOMICS INC.

XX (INCYTE GENOMICS INC.

XX (INCYTE GENOMICS INC.

XX (INCYTE GENOMICS INC.

XX (INCYTE GENOMICS INC.

XX (INCYTE GENOMICS INC.

XX (INCYTE GENOMICS INC.

XX (INCYTE GENOMICS INC.

XX (INCYTE GENOMICS INC.

XX (INCYTE GENOMICS INC.

XX (INCYTE GENOMICS INC.

XX (INCYTE GENOMICS INC.

XX (INCYTE GENOMICS INC.

XX (INCYTE GENOMICS INC.

XX (INCYTE GENOMICS INC.

XX (INCYTE GENOMICS INC.

XX (INCYTE GENOMICS INC.

XX (INCYTE GENOMICS INC.

XX (INCYTE GENOMICS INC.

CC gastrointestinal disorders (e.g. Crohn's disease), renal disorders (e.g. Good pasture's syndrome), viral, bacterial, fungal, parasitic, CC protozoal and helminthic infections, cardiovascular disorders (e.g. atherosclerosis), or hepatic diseases (e.g. cirrhosis) and many CC other diseases and disorders detailed in the specification. They can also CC be used in assessing the effects of exogenous compounds on the CC expression of nucleic acid and amino acid sequences of transporters and CC ion channels. TRICH or its fragments may also be used in screening for CC compounds that specifically bind to and modulate the activity of TRICH. CC The polynucleotides can be used to create knock-in humanised animals or CC transgenic animals to model human disease. The present sequence CC represents a TRICH protein.

SQ Sequence 374 AA:

Query Match 55.9%; Score 1961; DB 23; Length 374;

Best Local Similarity 99.7%; Pred No. 3e-192; Indels 0; Gaps 0;

Matches 373; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 300 MWYFTTAIGYPCPRYSNPADFYVDLTSIDRSRREDELATREKASIALFLEKVDLDF 359

DB 1 MWYFTTAIGYPCPRYSNPADFYVDLTSIDRSRREDELATREKASIALFLEKVDLDF 60

QY 360 LKRAETKLDDETCYESSVTPDTRNCLSPTRKMPGAVOQFTTLIRQISNDFRDLPTLLI 419

DB 61 LKRAETKLDDETCYESSVTPDTRNCLSPTRKMPGAVOQFTTLIRQISNDFRDLPTLLI 120

QY 420 HGAECALMSMTIGLFGHSGISLSPMDTALLFMIGALIPENYLDVYSQSPRAVLY 479

DB 121 HGAECALMSMTIGLFGHSGISLSPMDTALLFMIGALIPENYLDVYSQSPRAVLY 180

QY 480 YELEDGLYTTGTPYEFKILGELPEHCAYIIYGMPTVLANLRGLOPFLHFLVWLVY 539

DB 181 YELEDGLYTTGTPYEFKILGELPEHCAYIIYGMPTVLANLRGLOPFLHFLVWLVY 240

QY 540 FCCRIMAAALALPTFMASFNSALYNSYTLAGEFINUSLMTVPAMISKVSFLRWC 599

DB 241 FCCRIMAAALALPTFMASFNSALYNSYTLAGEFINUSLMTVPAMISKVSFLRWC 300

QY 600 FEGLMKIOFSRRTKMPGNLTIVASGDKILSAMELDSYPLATYLVIGSGFPVLYX 659

DB 301 FEGLMKIOFSRRTKMPGNLTIVASGDKILSAMELDSYPLATYLVIGSGFPVLYX 360

QY 660 VSLRFIRKQPSQDW 673

DB 361 VSLRFIRKQPSQDW 374

RESULT 5

AAG18079

ID AAG18079 standard; Protein: 632 AA.

AC AAG18079;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 19344.

CC Protein identification: signal transduction pathway; metabolic pathway; CC hybridisation assay; genetic mapping; gene expression control; promoter; CC termination sequence.

CC Arabidopsis thaliana.

CC EP1033405-A2.

CC 06-SEP-2000.

CC 25-FEB-2000; 2000EP-0301439.

CC 25-FEB-1999; 99US-0121825.

CC 05-MAR-1999; 99US-0123180.

CC 09-MAR-1999; 99US-0123548.

PR 22-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0138119.
PR 16-JUN-1999; 99US-0138452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0140333.
PR 23-JUN-1999; 99US-0140334.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140699.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152362.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.

```

PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 20.8%; Score 730.5; DB 21; Length 632;
 Best Local Similarity 30.7%; Pred. No. 2e-65;
 Matches 211; Conservative 117; Mismatches 269; Indels 91; Gaps 19;

```

OY 9 RGLKRGATPOTSGLODRLESSES--DNSLYFTSGCPNTLEVDLNTQVDLACQVWFE 66
DB 3 QGLDPMSTQSKSVLAPPTTISQGLQMSY-----PITKFEVYVKVI-----E 49
OY 67 QLAOFKMPWTPSPSCNCELGIQMLSPKVRSGOMLAIIGSSGCRASLLDVITGRGHGK 126
DB 50 QTSOCMSMKSE-----KTLNGITGVCPCGEFLAMGPGSGKTILLSLAGR--LSK 102
OY 127 IKSGQIWMGQSSSPQVLKRCVAHVRQHNQLLPMLVRETLAFLAQNMLPRTFSSOARDK 186
DB 103 TFSCKVMYNGQPFSGCIKRR-TGFVADDDVLYPHLFWETLEFALLRLPSTLRDEKAE 161
OY 187 RVEDVIAELRLQCADRVGMATYRGLSGERRRVSIGVOLLMPGIIIDEPISGDSF 246
DB 162 HDNRVIAELGLMRCNTSMIGPLRGISGGEKKRVSIGQEMLIMPSTLLIDEPISGDSF 221
OY 247 TANHVLVTLISRLAGNRLVLSLHQPSPDIFRLFDVLVMTSGPTVIGAAQHMVOYETA 306
DB 222 TANHIVTTIKRLAAGGRVYTTIHOPSRITVHMDKVLSESGPTIYGAASSAVEYFSS 281
OY 307 IGYCPKRYSNADFYVDLTS-----IDRSREDELATREKQSLAFLKVRDLDFLW 361
DB 282 LGFSTSLVNPADLLDLANGIIPDQKETSEQOKTVK--ETLYSAVEKNI----- 331
OY 362 KAEKRLDEDCVSSS---VTPIDTNCLSPTMPAQQVFTLLIRQI-SNDFRLDPT 416
DB 332 --STK-LKAEICNMSHSYETKAAARNLSEOMCTWVQFVLLRQVARERFESFNK 388
OY 417 LLIHGAECALMSMTIGFLYFGHSIQLSFMDTALLFMIGALLPFNVIIIDVISCYSERA 476
DB 389 LRIF---QVISVALIGLILMWH-TPKSHIDRRLALLFFSVFMCFTLYLVAVPTFPDEKR 444
OY 477 MLYTELEDGLYTTGPFYFANLIGELPERCAVYIIYIGMPTVLANLRGLOPFLILFLVW 536
DB 445 MLIKERSSGMYRLSSYFMAWVGDLPLELAPFAVFIYIWMGCLKRDPFTFILLSLVYL 504
OY 537 LVYCCNIMLALALALPFTFHMASFSPNALXNSYLAGGFMINISLMTVP---AMISKV 593

```

```

DB 505 YSLVAOGLGLAFGLLNNIKQATTLASVTLVFLIAGGYVQ-----QIPPTVWAKYL 559
OY 594 SELRCFEGJLMIKIOFSRRTY-----KMPIGNLTTAVSGDKTILSNMEL 635
DB 560 SYSTYCYKLLGLGIOTDDYIEYCSGWACRGDPPALIKSMGLNNIMT---DVEFMGYML 615
OY 636 DSYPLAIVLYIGISGGFVNLVYVSLR 663
DB 616 VGYRLMA-----YMLLHVYKLR 632

```

RESULT 6
 ID AAG18078 standard; Protein; 648 AA.

AC AAG18078;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 19343.

KW Protein identification; signal transduction pathway; metabolic pathway;
 hydridisation assay; genetic mapping; gene expression control; promoter;
 termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

```

XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.

```

PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140053.
PR 23-JUN-1999; 99US-0140054.
PR 24-JUN-1999; 99US-0140055.
PR 24-JUN-1999; 99US-0140056.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0143247.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145222.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148664.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 20.8%; Score 730.5; DB 21; Length 648;
Best Local Similarity 30.7%; Pred. No. 2,1e-65;
Matches 211; Conservative 117; Mismatches 269; Indels 91; Gaps 19;

PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145216.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148177.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151067.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match

Best Local Similarity 30.9%; Score 724; DB 21; Length 625;
Matches 211; Conservative 113; Mismatches 258; Indels 100; Gaps 20;

QY 15 ATPQDTS--GADRLFSSESNSLYFTYSCGPNLEVDLNVQVDLASQVWFEOLOAFK 72
DB 11 APTFTSQPGLQ-----MSM-----PITLKEFEVYKTKI-----EOTSOM 48
QY 73 MPWTSPSCQNCCEJGIONLSTFRVSGOMLAIIGSGCGRASLDVYTGGRGKIKSGOI 132
DB 49 GSKMSKE-----KTLINGITGMVCBEFLAMGPGSGKTTLLSMGCR--LSKTFSGKV 101
QY 133 WINGOPSSPOLVRVCVANHVRHNDLPLVTRFETLAFIAQNRLEPRTSQAORDKRVDEVI 192
DB 102 MYNQPFSGCIRKR--TGFEVADDDVLYPHLVTWETLFEFALLRLPSSILRDEKAEHVRI 160
QY 193 AELRLROCADTRVGNMVRGLSGGERRRVSIGVOLLNPGILIDEPTSGDSFTAHNLV 252
DB 161 AELGLNCTNMIGGPLERFSGGEKRVISIGOEMLNPSLLIDEPTSGDSTTAHRIV 220
QY 253 KTLRLAKNRLVLIISHQPSDIJRLFDVLVLMTSGTPIYGAQHNVQYFTAGYPCP 312
DB 221 TTIRKLASGGRTVVTYTHIOPSSRIYHMDKVLLSEGPSIYYGAASSAVEYFSSIGSTS 280
QY 313 RYSNPADFYVDLTS-----IDRSRQELATREKAQSIALFLEKVRDLDDFLKAETKD 367
DB 281 LTVNPADLLDLANGIPDPDROKTESQEOQKVK--ETLVSAYEKNI-----STK- 327
QY 368 LDEDTCVES--VFPLDTNCLPSPTKMPGAVOOFLLHROI--SUDFRLDPTLLHGA 422
DB 328 LKAEICNAESHSTYTTAAKANKLSBOMCTTWTQFVLQGVKRRERFSPFKLRFP-- 385
QY 423 EACLSMTIGFLYFGHCSIQLSFMDTAALEFMICALIPFNVLIDVISCYSEKRALYYEL 482
DB 386 -QVTSVAFILGLIMWH--TPKSHIDRTALLFEFFSVGFPLVNAVTFPOEKRMILKER 443
QY 483 EDGITYTGPRYFAKILDELBHCYIITIGMPYTWLANLRGLOPFLHFLVLVAVVCC 542
DB 444 SSGMYRLSYFEMANNVDLPLALPRAVFIIYIMGGLKRPDDPTFTLSLVVLYSVLVA 503

CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 protein of the invention. This
 CC sequence is encoded by the human ABCG5 gene located on chromosome 2p21.
 XX

Sequence 651 AA;

Query Match 19.9%; Score 697; DB 23; Length 651;
 Best Local Similarity 28.9%; Pred. No. 5,9e-62;

Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;

16 TPQDTSGLDRLFSSESNSLYFTYSGOPTLEVRDNLVQVDAASQVPEFOLAQFKMPW 75
 8 TPGGSMGLQVNRGSSGLEGAPAT-APRPHSLGILHASYSHRVR-PWMD-ITSCROOW 64
 76 TSPSCONSCELGIONLSFKVRSQOMLAIIGSSCGRASLDVITGR-GHGKIRKSGQIWI 134
 65 TROI-----LKDVSILYVESGQIMCIGSSGCKTLLDAMSRLGRAGTF-LGEYVY 115
 135 NGQFSSPOLVRKCYAHVROHNOILPNLTVRETLAFIAQMRKLPFRFSQARKRVEDVYAE 194
 116 NGRARRRQFQDCSYVQSTPLSLTLVRETLHTALAI-RRGNPQSPFKVEAVVAE 174
 195 LRLROCADTRVGNMVRNGLSGERRRVSIGVOLLMPGILLDEPTSGLDSEFTAHNYKT 254
 175 LLSLVADRLRIGNSISGIRERRRSIAQLODPRVMLFDEPTTGLDCTANQIYVL 234
 255 LSLAKGNRLVLSLHOPRSDIFRLFDVLLMTSGTPYIIGAOMHVOYFPAAGPCRY 314
 235 LVELARRRRIYVLIHOPRSELFDLFDIALISFCELLFCGPAEMLDFNDGCPPEH 294
 315 SNPDEFVLDLSTIRSRREOLATREKAOISALF-----LEKRDLDPLMK 362
 295 SNPDEFVLDLSTIRSRREOLATREKAOISALF-----LEKRDLDPLMK 362
 363 AETKDLDEDCVSSVPLDTNCLPSPK-MPGAVQOFTLLIRQISNDFDLPTLLIHG 421
 349 -----KTLPM-----VPRKDKSPGVFSKLGVLRRVTRNLVANKLAVITRL 390
 422 AEACLMSTIGFLYFG-----HSTQLSEMDTALLFMIGALIPFNVILVYISKYSER 475
 391 LQNLIMGLFLFVLVRSNVLKGAIO-----DRVGLLYQFVGATPYGMLNVAVLPVLR 446
 476 AMLYELEDGLYTTGPFPAKILGELPEHCAYIIITYGMPTWLANLRGQPLHLHLV 535
 447 AVSQOESQDGLYQKQOMLAVLALVLPSSVAVMIFSSVCTGLLPEVARF----- 499
 536 WLVECCRIMALAALLPTFHMASFVS-----NALYSEFYLAG-----GFM 577
 500 -----GYFAALLAPHLIGEFLLVGLVQNPNIYNSVALLSTAGVAGGFL 549
 578 INLSLMTVPAKIKSVFLKRCFGLMKIQPSRRTYMPGLNLTIAVS 625
 550 RNIDEMPIFRITISYFTFOKYSILVNEFYGLNFCGSSNVSVTN 597

RESULT 11

AAU96993 standard; Protein; 651 AA.

AAU96993;

30-JUL-2002 (first entry)

XX Human ABCG5 mutant R419p protein sequence.
 DE Human: ABCG5: ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 XX arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW mutant; mutelin.
 XX Homo sapiens.
 OS Synthetic.
 XX Key Location/Qualifiers
 FT Misc-difference 419
 FT /note="Wild-type Arg substituted by Pro"
 XX
 PN WO200227016-A2.
 PD 04-APR-2002.
 XX 25-SEP-2001; 2001WO-US29859.
 PF 25-SEP-2000; 2000US-235268P.
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 XX Patel SB, Dean M;
 PI WPI: 2002-416483/44.
 DR Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 PS
 XX Claim 10; Page -: 66pp; English.

The present invention relates to a new mammalian ATP-binding cassette
 gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 predisposition for developing sitosterolemia, arteriosclerosis or heart
 disease. The molecules of the invention are also useful for identifying
 a compound which alters ABCG5 activity level comprising contacting a cell
 culture or mammal which have ABCG5 polypeptide with a compound and
 measuring ABCG5 biological activity in the cell culture or in mammal,
 where an increase or decrease in ABCG5 biological activity compared to
 ABCG5 biological activity in a control cell culture or mammal not
 contacted with the compound, identifies a compound that increases or
 decreases ABCG5 activity respectively. The cell culture or mammal
 comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 mutant R419p protein of the
 CC invention.
 CC Note: This sequence is not shown in the specification but is derived
 CC from the wild-type human ABCG5 protein (AAU96984) given on pages 35-36
 CC of the specification.
 XX

Sequence 651 AA;

Query Match 19.9%; Score 697; DB 23; Length 651;
 Best Local Similarity 28.9%; Pred. No. 5,9e-62;

Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;

16 TPQDTSGLDRLFSSESNSLYFTYSGOPTLEVRDNLVQVDAASQVPEFOLAQFKMPW 75
 8 TPGGSMGLQVNRGSSGLEGAPAT-APRPHSLGILHASYSHRVR-PWMD-ITSCROOW 64
 76 TSPSCONSCELGIONLSFKVRSQOMLAIIGSSCGRASLDVITGR-GHGKIRKSGQIWI 134

Accession	Species	Gene	Protein	Position	Sequence	Position	Sequence
Db	65	THQI	-----LKVSIKVESQIMCILLSSSSCKTTLIDAMSGLRAGCT-LGEVY	113			
Oy	135	NGQPSSPQLVRCVAHVHROHNOILLPNLVRETIAFLTAQMKLPPTESQAOQRKREVDVIAE	194				
Db	116	NGRALRREOFQDCFSYVLQSDTLSSLVRETLHYTALLAI-RKNGPGRQKKEVAEMAE	174				
Oy	195	LRLRCQADPRVGNMVRGLSGGERRRVSIGVQLMMPGIILDEPFTSGDLSFPAHNLVKT	254				
Db	175	LSLSHVADRLIGNYISGIGISTGERRRVSIQAOLDQPKWMLDEPFTGLDCMTANOIVVL	234				
Oy	255	LSRLAKGNLVLSLHQPRSDIFRLDELVLMLTSGTPIYGAQAQHWQYFTATGCGCPRY	314				
Db	235	LYELARRNRIIVLTTHQPSSELFQFDKTAIISFGELICGTPAEMLDFPNDGYPCEP	294				
Oy	315	SNPAPDYVDLSIRDRSRQELATREKQSLAALF-----LEKVRDLDDFLWK	362				
Db	295	SNPFDFYMDLTSVDYQSKEREIETSKRVQMISAYKKSACHTKLNIERMKHL-----	348				
Oy	363	AETKRLDEPTCVESSTPLDTNCLSPSTR-MGAVAQFTTLIRROISNDPRLPVLING	421				
Db	349	-----KTLPM-----VPEKTDSPGVESKGLVLLRRVTRNLVRNKLAVITRL	390				
Oy	422	AEACLSMTIGFLYFG-----HGSITLSFMPTAALLFMIGALLPENNVILDIYSKYSRR	475				
Db	391	LQNLINGLEFLFFVLRKSNVNLKGAIO-----DPVGLLIQFVGATPPTGMNNAVNEPVL	446				
Oy	476	AMLYYELEDEGLYTPPYEFPAKILGELPEHCAYIIYGMPTVLMANLRPGLQFPLHFLLY	535				
Db	447	AVSDQESQDGLYQKQOMMLAYALHYLPFSVAVATMIFSSVCYWTGLGHPVARF-----	499				
Oy	536	WLNVCCRIMLAALALPFTFHMASFFS-----NALINSEYLAG-----GFM	577				
Db	500	-----GYFSALLAPHLIGELFLTVLGIYQNPVIVNSVALLSIAGVLVSGSFL	549				
Oy	578	INLSLMTVPAMISKVSFLRMCFEGELMKIOFSRRTYKMKLGLTIAVS	625				
Db	550	RNIQEMPIPKIISITFTQKTCSELTIVNEFYGLNFTGSSNSVATTN	597				
RESULT 12							
AAE13290							
ID	AAE13290	standard; Protein; 651 AA.					
XX	AAE13290;						
DE	12-FEB-2002	(first entry)					
KW	Human	sticosterolaemia susceptibility gene (SSG) protein.					
KM	Human; sticosterolaemia susceptibility gene; SSG; atherosclerosis;						
KV	steroid-related disorder; hyperlipidaemia; hypercholesterolaemia; therapy;						
KX	gall stone; coronary heart disease; cardiovascular disease; arthritis;						
XX	xanthoma; haemolytic anaemia; transgenic animal; chromosome 2p21.						
OS	Homo sapiens.						
XX	MO200179272-A2.						
XX	25-OCT-2001.						
XX	18-APR-2001; 2001WO-US12758.						
XX	18-APR-2000; 2000US-198465P.						
XX	15-MAY-2000; 2000US-204234P.						
PA	(TULA-) TULARIK INC.						
PI	Tian H, Schultz J, Shan B;						
XX	WPI; 2002-017598/02.						
DR	N-PSDB; AAD22009.						
XX	Novel sticosterolaemia susceptibility gene polypeptide and						
XX							

us-09-989-981a-7.rng

Page 1

OM nucleic - nucleic search, using sw model

Run on: July 26, 2003, 17:02:14 ; Search time 645.803 Seconds

(without alignments) :
1156.341 Million cell updates/sec

Title: US-09-989-981A-7

Perfect score: 2669
Sequence: 1 ggtgtccctgtctccaggaac.....caattaaaatgtattgagc 2669

Scoring table: IDENTITY_NUC

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 08
Maximum Match 1000

Listing first 45 summaries

Database : N_Geneseq_19Jun03:*.

1:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA1980.DAT.*
2:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA1981.DAT.*
3:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA1982.DAT.*
4:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA1983.DAT.*
5:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA1984.DAT.*
6:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA1985.DAT.*
7:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA1986.DAT.*
8:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA1987.DAT.*
9:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA1988.DAT.*
10:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA1989.DAT.*
11:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA1990.DAT.*
12:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA1991.DAT.*
13:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA1992.DAT.*
14:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA1993.DAT.*
15:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA1994.DAT.*
16:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA1995.DAT.*
17:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA1996.DAT.*
18:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA1997.DAT.*
19:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA1998.DAT.*
20:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA1999.DAT.*
21:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA2000.DAT.*
22:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA2001.DAT.*
23:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA2001B.DAT.*
24:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA2002.DAT.*
25:	/SIDSL/gcgdata/genseq/genseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2669	100.0	2669	25	AAD48883	Human- ABCG8 DNA.
2	1680.6	63.0	3239	24	ABK83218	Human transporter
3	1499	56.2	2564	24	ABN90022	Mouse clone IMX3_6
4	1430	53.6	2019	25	AAD4881	Mouse ABCG8 DNA.
5	291.6	10.9	580	22	AAH98911	Arabidopsis EST-depS
6	203.6	7.6	1920	24	AAK51681	DNA encoding human
7	203.6	7.6	2340	24	AAD22009	Human ABCG5
8	203.6	7.6	2340	25	AAD48882	Human ABCG5 DNA.

[illegible]

ALIGNMENTS

RESULT 1
AAD48883

ID.	standard; DNA; 2669 BP
xx	

AC AAD48883;

DT 24-MAR-2003 (first entry,
yy

Human ABCG8 DNA.

KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;

HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy

human, HIV-binding cassette; sirosterolaemia susceptibility gene; SSG, ABCG5: gene; ds

XX
XX
XX

XX

FT	CDS	100..2121

```

/*lag= a
/ord= a
#harcos state=1
E1
E1

```

XX
XX

XX

XX

20-NOV-2001; 2001WU-US43823

PR 20-NOV-2000; 2000US-252235P

XX (TULAR) TULARIK INC.
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Hobbs HH, Shan B, Barnes R, Tian H;
XX MPI: 2003-058548/05.
DR P-PSDB: AAE31705.
XX
PT New ABCG8 polypeptides and nucleic acids, useful for treating
PT sterol-related disorders e.g. sitosterolemia, hypercholesterolemia,
PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
PT nutritional deficiencies -
XX
PS Claim 13; Page 80; 94pp; English.
XX
XX The invention relates to ATP-binding cassette (ABC) family cholesterol
CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also
CC known as sitosterolemia susceptibility gene (SSG). Sequences of the
CC invention are useful for treating or preventing sterol-related disorders
CC such as sitosterolemia, hyperlipidemia, hypercholesterolemia, gall
CC stones, HDL deficiency, atherosclerosis and nutritional deficiencies.
CC They are also useful in gene therapy. The present sequence is human
CC ABCG8 DNA.
XX
XX Sequence 2669 BP; 595 A; 768 C; 722 G; 584 T; 0 other:
SO
Query Match 100.0%; Score:2669; DB 25; Length 2669;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2669; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GTGTCCCTGCTCCAGAAACAGAGTGAAGACATGCGCCCTGGACAGCAGCTGGGTCT 60
DB 1 GTGTCCCTGCTCCAGAAACAGAGTGAAGACATGCGCCCTGGACAGCAGCTGGGTCT 60
OY 61 AAGAGAGTGCAGGCCGAGGCTCAGAGACTGTGGGCGCCGATGCGCGGAGCGGACAG 120
DB 61 AAGAGAGTGCAGGCCGAGGCTCAGAGACTGTGGGCGCCGATGCGCGGAGCGGACAG 120
OY 121 GAGAGAGGCTGCCGAAAGGGGCGACCTCCAGAGTACCTGGGCTCCAGATAGATTG 180
DB 121 GAGAGAGGCTGCCGAAAGGGGCGACCTCCAGAGTACCTGGGCTCCAGATAGATTG 180
OY 181 TTCTCTCTGAAATGACAAAGCTGTACTTCACTACATGAGTGGCCAGCCACACCTG 240
DB 181 TTCTCTCTGAAATGACAAAGCTGTACTTCACTACATGAGTGGCCAGCCACACCTG 240
OY 241 GAGGTGAGAGACCTCACTACAGTGGAGTGGGCTGCTGAGTGGCTGGTGGTGGAG 300
DB 241 GAGGTGAGAGACCTCACTACAGTGGAGTGGGCTGCTGAGTGGCTGGTGGTGGAG 300
OY 301 CTGGCTCAGTTCAAGATGCCCTGGACATCTCCAGCTGCGCAGATTTCTTGAGCTGG 360
DB 301 CTGGCTCAGTTCAAGATGCCCTGGACATCTCCAGCTGCGCAGATTTCTTGAGCTGG 360
OY 361 ATTCAGAACTTAAGCTTCAAGTGAAGTGGGAGATGCTGGCCATATAGGAGCTCA 420
DB 361 ATTCAGAACTTAAGCTTCAAGTGAAGTGGGAGATGCTGGCCATATAGGAGCTCA 420
OY 421 GGTGTGGAGAGGCTCTGCTAGATGATCACTGCGGAGGTCACGGCGGCAATC 480
DB 421 GGTGTGGAGAGGCTCTGCTAGATGATCACTGCGGAGGTCACGGCGGCAATC 480
OY 481 AAGTCAGGCCAGATCTGATCAATGGGAGCCGAGCTCGCTCAGCTGGTGAAGAAATGT 540
DB 481 AAGTCAGGCCAGATCTGATCAATGGGAGCCGAGCTCGCTCAGCTGGTGAAGAAATGT 540
OY 541 GTGGCCACGTTGGGCGGACACACAGCTGCTCCCAACTTGAATGACAGAGACCTTG 600
DB 541 GTGGCCACGTTGGGCGGACACACAGCTGCTCCCAACTTGAATGACAGAGACCTTG 600
OY 601 GCCTTCATGCTCCAGATGCGGCTGCGCAGAAACCTTCTCCAGGCGCCAGGCTGACAAAG 660

DB 601 GCCTTCATGCTCCAGATGCGGCTGCGCAGAAACCTTCTCCAGGCGCCAGGCTGACAAAG 660
OY 661 GTGAGAGAGATGAGGAGGAGCTGGGCTTGGAGCAGTGGCGTCAACCGGCTGGGCAAC 720
DB 661 GTGAGAGAGATGAGGAGGAGCTGGGCTTGGAGCAGTGGCGTCAACCGGCTGGGCAAC 720
OY 721 ATGTACGTGGGAGGTTGTGGGAGGAGGAGGAGAGTACATTTGGGTGGAGCTC 780
DB 721 ATGTACGTGGGAGGTTGTGGGAGGAGGAGGAGAGTACATTTGGGTGGAGCTC 780
OY 781 CTGTGGAACCCAGAGATCTTATCTGACAGAACCCACTCTGGGCTGACAGCTTCA 840
DB 781 CTGTGGAACCCAGAGATCTTATCTGACAGAACCCACTCTGGGCTGACAGCTTCA 840
OY 841 GCCCACAACCTGTTGAAGACCTTGTCCAGGCTGGGCGCAAGGCAACCGGCTGGTCA 900
DB 841 GCCCACAACCTGTTGAAGACCTTGTCCAGGCTGGGCGCAAGGCAACCGGCTGGTCA 900
OY 901 TCCCTCCACAGCCTCGCTGACATCTTACAGGCTGTTGATGCTGCTCTGATGACG 960
DB 901 TCCCTCCACAGCCTCGCTGACATCTTACAGGCTGTTGATGCTGCTCTGATGACG 960
OY 961 TCTGGCACCCCACTACTAGTGGGCGGCGCCAGACATGCTCAGTATTTCAAGCCATC 1020
DB 961 TCTGGCACCCCACTACTAGTGGGCGGCGCCAGACATGCTCAGTATTTCAAGCCATC 1020
OY 1021 GGCTAACCCCTGCTCTGCTAGACATCTCTGACTTCTATGAGCTTACCAAGATT 1080
DB 1021 GGCTAACCCCTGCTCTGCTAGACATCTCTGACTTCTATGAGCTTACCAAGATT 1080
OY 1081 GACAGGCGCAGACAGAGAGAGAAATGGCCACAGGAGGAGAGGCTCAGTCACTGCA 1140
DB 1081 GACAGGCGCAGACAGAGAGAGAAATGGCCACAGGAGGAGAGGCTCAGTCACTGCA 1140
OY 1141 CTGTTTCTAGAAAAGTGTGCTGACTTATGATGACTTTTATGGAAGAGAGACAGAG 1200
DB 1141 CTGTTTCTAGAAAAGTGTGCTGACTTATGATGACTTTTATGGAAGAGAGACAGAG 1200
OY 1201 CTTGACGAGGACACTGTGTGGAAGAGAGGCTGACCCCACTAGACACCAACTGCTCC 1260
DB 1201 CTTGACGAGGACACTGTGTGGAAGAGAGGCTGACCCCACTAGACACCAACTGCTCC 1260
OY 1261 AGTCTACGAAAGATGCTGGGCGGTGTCAGAGTTCAGAGCTGATCCGTCAGATT 1320
DB 1261 AGTCTACGAAAGATGCTGGGCGGTGTCAGAGTTCAGAGCTGATCCGTCAGATT 1320
OY 1321 TCACAAGACTTCCGAGACCTGCGCACCCTCTCATTCATGAGGCGGAGGCTGTCTG 1380
DB 1321 TCACAAGACTTCCGAGACCTGCGCACCCTCTCATTCATGAGGCGGAGGCTGTCTG 1380
OY 1381 TCATGACCAATCGGCTCTCTATTTTGGGCAATGGGAGCAATCCAGCTCTCTCATG 1440
DB 1381 TCATGACCAATCGGCTCTCTATTTTGGGCAATGGGAGCAATCCAGCTCTCTCATG 1440
OY 1441 ACAGCCGCCCTCTTGTTCATGATCGTGTCTCATCCCTTCAACGCTATTTGGATG 1500
DB 1441 ACAGCCGCCCTCTTGTTCATGATCGTGTCTCATCCCTTCAACGCTATTTGGATG 1500
OY 1501 ATCTCCAAATGTTACTAGAGAGGCGCAATCTTACTATGAACTGGAAGAGGGCTGTAC 1560
DB 1501 ATCTCCAAATGTTACTAGAGAGGCGCAATCTTACTATGAACTGGAAGAGGGCTGTAC 1560
OY 1561 ACCACGTGTCATATTTCTTGGCAAGATCTCTGGGAGAGTTCGCGAGCACTGTGCTAC 1620
DB 1561 ACCACGTGTCATATTTCTTGGCAAGATCTCTGGGAGAGTTCGCGAGCACTGTGCTAC 1620
OY 1621 ATCATCATCTAGAGGATGCCACCTACTGCTGCGCAACCTGAGAGGCGCTCCAGCCC 1680
DB 1621 ATCATCATCTAGAGGATGCCACCTACTGCTGCGCAACCTGAGAGGCGCTCCAGCCC 1680
OY 1681 TTCTCTGTGACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740

Db	1681	TTTCGCTGACACTTCTCTGTGGTGTGGCTGGTGGTCTTCTGTGTGACGATTATATGGCCCTG	1740
OY	1741	GGCGCGCGG6CCCTGCTCCCACTTCCACATG6CCTCTCTTTCAGCAATG6CCCTTAC	1800
Db	1741	GGCGCGCGG6CCCTGCTCCCACTTCCACATG6CCTCTCTTTCAGCAATG6CCCTTAC	1800
OY	1801	AACCTCTTCTACCTGCGCGGGGGGCTTCATGATTAACCTTGAGCAGCCTGTGACAGTGGCC	1860
Db	1801	AACCTCTTCTACCTGCGCGGGGGGCTTCATGATTAACCTTGAGCAGCCTGTGACAGTGGCC	1860
OY	1861	GGCTGATTTTCCAAAGTGTCCCTTCTCGGGGTGTGTTTGAAGAGGCTGATGAATTTACG	1920
Db	1861	GGCTGATTTTCCAAAGTGTCCCTTCTCGGGGTGTGTTTGAAGAGGCTGATGAATTTACG	1920
OY	1921	TTTCACGACAGAACTTATTAATATGCTCTCGGGAACCTCACATCGCGGTCTCAGAGAT	1980
Db	1921	TTTCACGACAGAACTTATTAATATGCTCTCGGGAACCTCACATCGCGGTCTCAGAGAT	1980
OY	1981	AAAAATCCAGAGCCATGAGCTGAGCTGACGCTGATACCTCTCTACGCACTTACTCTATGCTC	2040
Db	1981	AAAAATCCAGAGCCATGAGCTGAGCTGACGCTGATACCTCTCTACGCACTTACTCTATGCTC	2040
OY	2041	ATTGGCCTCAGCGGTGCTTATGCTCTGATCTACTACGCTGCTTAAGTTATATCAACAG	2100
Db	2041	ATTGGCCTCAGCGGTGCTTATGCTCTGATCTACTACGCTGCTTAAGTTATATCAACAG	2100
OY	2101	AAACCAAGTCAAGACTGGTGTGATTACAGCCAGACGCTG6CCGCTGTGTGGGGAGCCTGAGC	2160
Db	2101	AAACCAAGTCAAGACTGGTGTGATTACAGCCAGACGCTG6CCGCTGTGTGGGGAGCCTGAGC	2160
OY	2161	AGACCCCTTCAACTGCATCCTCTCTCAGAGAACCCCTTCTGTGGGACAGTGAAGCAATGA	2220
Db	2161	AGACCCCTTCAACTGCATCCTCTCTCAGAGAACCCCTTCTGTGGGACAGTGAAGCAATGA	2220
OY	2221	CCCTACAGATCTGAGTACATTCGCGGCCAGAGGTGTGACATGTGGCACAGACACCACAG	2280
Db	2221	CCCTACAGATCTGAGTACATTCGCGGCCAGAGGTGTGACATGTGGCACAGACACCACAG	2280
OY	2281	GATGGCAGTAGAATTAAGACAGTGTGAAAGGATTTGCTGCTACCTGAGCAGAGACTGGCAT	2340
Db	2281	GATGGCAGTAGAATTAAGACAGTGTGAAAGGATTTGCTGCTACCTGAGCAGAGACTGGCAT	2340
OY	2341	GACTGGGAGAAACCTGCACTCGGTGGACCACTACAAAGTGTCTAATTTATTTCTTTTGA	2400
Db	2341	GACTGGGAGAAACCTGCACTCGGTGGACCACTACAAAGTGTCTAATTTATTTCTTTTGA	2400
OY	2401	TATGCATTTTATATGGCACTGCACTGATATATAGATGTGGAGCAAACTATGATATTTGGGTAG	2460
Db	2401	TATGCATTTTATATGGCACTGCACTGATATATAGATGTGGAGCAAACTATGATATTTGGGTAG	2460
OY	2461	CTAGACTGTGCAGAAATTTGTGAAACCTGTGAGGGAGAACATTAACAGTAGTAGAGATTGG	2520
Db	2461	CTAGACTGTGCAGAAATTTGTGAAACCTGTGAGGGAGAACATTAACAGTAGTAGAGATTGG	2520
OY	2521	GCTTCATCTTCCAGGGGCCCCACACTCCGTGTGAGCCACCATCAATATACAGAAAGTGACC	2580
Db	2521	GCTTCATCTTCCAGGGGCCCCACACTCCGTGTGAGCCACCATCAATATACAGAAAGTGACC	2580
OY	2581	TAAATATTACACAGCAATGCACTCTCTTTTGTGTGTGGGTGATGAGGCTCCAAAGC	2640
Db	2581	TAAATATTACACAGCAATGCACTCTCTTTTGTGTGTGGGTGATGAGGCTCCAAAGC	2640
OY	2641	CAACGTGAACAATTAATAATGTATTTGAGC	2669
Db	2641	CAACGTGAACAATTAATAATGTATTTGAGC	2669

27-AUG-2002 (first entry)

Human transporter and ion channel, TRICH9, Incyte ID 6585710C81, CDNA.

Human; ss; gene; transporter and ion channel; TRICH; transport disorder; neurological disorder; muscle disorder; immunological disorder; cancer; scleroderma; systemic lupus erythematosus; allergy; leukaemia; cell proliferative disorder; cervical cancer; breast cancer; neurodegenerative disorder; Parkinson's disease; Alzheimer's disease; myotonic dystrophy; catatonia; endocrine disorder; diabetes; Grave's disease; gastrointestinal disorder; Crohn's disease; renal disorder; good pasture's syndrome; viral infection; cirrhosis; bacterial infection; fungal infection; parasitic infection; protozoal infection; helminthic infection; parasitic infection; atherosclerosis; hepatic disease.

Homo sapiens.

WO200240541-A2.

23-MAY-2002.

25-OCT-2001; 2001WO-US46055.

27-OCT-2000; 2000US-243989P.
03-NOV-2000; 2000US-245904P.
09-NOV-2000; 2000US-247673P.
17-NOV-2000; 2000US-249661P.
20-NOV-2000; 2000US-252232P.
01-DEC-2000; 2000US-250790P.

(INCYTE GENOMICS INC.

Tang YT, Yue H, Nguyen DB, Hafalia AJA, Elliott VS, Lu Y; Walla NK, Yao MG, Baughn MR, Gandhi AR, Ding L, Sanjanwala M; Rankumar J, Arvizu C, Gietzen KT, Lal PG, Azimzal Y, Khan FA; Thangavelu K, Thornton M, Lu DM, Tribouley CM, Warren BA; Ison CH, Das D, Raumann BE, Policky JL, Kearney L; WPI; 2002-463570/49.
P-PSDB; ABG61539.

New transporters and ion channels (TRICH) polypeptides, useful for diagnosing, preventing, and treating disorders associated with an abnormal expression or activity of TRICH, e.g. immunological, muscular or renal disorders -

Claim 5; Page 167-168; 178pp; English.

The invention relates to human transporters and ion channels (TRICH) polypeptides, a naturally occurring amino acid sequence 90 % identical to TRICH, a biologically active fragment of TRICH or an immunogenic fragment of TRICH. Also included are an isolated polynucleotide encoding TRICH, a recombinant polynucleotide comprising a promoter sequence operably linked to the TRICH polynucleotide, a cell transformed with the recombinant polynucleotide, a transgenic organism comprising the recombinant polynucleotide, an isolated antibody that binds specifically to TRICH, and screening for compounds which bind to TRICH, modulate TRICH, modulate TRICH expression or are ant/agonists of TRICH. The polypeptides are useful for diagnosing, preventing, and preventing transport, neurological, muscle, immunological disorders (e.g. scleroderma, systemic lupus erythematosus, allergies), cell proliferative disorders such as cancers (e.g. leukaemia, cervical or breast cancers), neurodegenerative disorders (e.g. Parkinson's disease, Alzheimer's disease), muscular disorders (e.g. myotonic dystrophy, catatonia), endocrine disorders (e.g. diabetes, Grave's disease), gastrointestinal disorders (e.g. Crohn's disease), renal disorders (e.g. Good pasture's syndrome), viral, bacterial, fungal, parasitic, protozoal and helminthic infections, cardiovascular disorders (e.g. atherosclerosis), or hepatic diseases (e.g. cirrhosis) and many other diseases and disorders detailed in the specification. They can also be used in assessing the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of transporters and

Db 852 GCCTCATGATAAAGCTTGACAGCCCTGAGACATGGCCGGCTGGATTCCCAAGTGTCTT 911

Oy 1883 TCCTGGGTGGTGTTTTGAAGGCGCTGATAGATTACGTTCCACAGAGAAGACTTATPAAA 1942

Accession	Gene	Position	Sequence	Length
Dp		912	TCCGCGGTGGTGTGTTTGAAGGGCTGATGGAAGTTCAGTTTCAGAGAAAGAACTTATATAA	971
Qy		1943	TGCGCTCTCGGGAACTTACCATGCGGGTCTCAGAGATATAAATCTCAGTGCATGGAGC	2002
Dp		972	TGCGCTCTCGGGAACTTACCATGCGGGTCTCAGAGATATAAATCTCAGTGCATGGAGC	1031
Qy		2003	TGAGCTGTACCCCTCTTACGCCATATACCTCATTCGATTTGGCTGAGGGTGGCTTCA	2082
Dp		1032	TGAGCTGTACCCCTCTTACGCCATATACCTCATTCGATTTGGCTGAGGGTGGCTTCA	1091
Qy		2063	TGCTCTGTACTACGTGTCTTAAGTTTCATCAACGAAGAACCAAGTCAAGACTGTGAT	2122
Dp		1092	TGCTCTGTACTACGTGTCTTAAGTTTCATCAACGAAGAACCAAGTCAAGACTGTGAT	1151
Qy		2123	TCACGCCAGACCTCTGCGCCCTGTGGGGACCTGACAGACCCCTTCACTGCATCCT	2182
Dp		1152	TCACGCCAGACCTCTGCGCCCTGTGGGGACCTGACAGACCCCTTCACTGCATCCT	1211
Qy		2183	CCTCAGAGAGCCCTCTTCTGGGGACAGTGAAGACATGACCTCAGATGGTCACTCAT	2242
Dp		1212	CCTCAGAGAGCCCTCTTCTGGGGACAGTGAAGACATGACCTCAGATGGTCACTCAT	1271
Qy		2243	CCGGCCAGGGGTCTCAGTGGCAGACAGACGACAGATGGCAGTGAATTAAGACAG	2302
Dp		1272	CCGGCCAGGGGTCTCAGTGGCAGACAGACGACAGATGGCAGTGAATTAAGACAG	1331
Qy		2303	TCGAAAGGATTTCTGCTCAGTGGCAGACGACGATGATCTGGGAGAAAACCTGCATC	2362
Dp		1332	TCGAAAGGATTTCTGCTCAGTGGCAGACGACGATGATCTGGGAGAAAACCTGCATC	1391
Qy		2363	GGTGGACCTACACGTTGCTAATTTATTTCTTTTGTATGATTTATATATAGCACTC	2422
Dp		1392	GGTGGACCTACACGTTGCTAATTTATTTCTTTTGTATGATTTATATATAGCACTC	1451
Qy		2423	GATATGAGTGGAGAGCAATAGAGATGATTTGGTATGCTTACACTGTGACGAAATGTTG	2482
Dp		1452	GATATGAGTGGAGAGCAATAGAGATGATTTGGTATGCTTACACTGTGACGAAATGTTG	1511
Qy		2483	GAACTGGAAGGAGCAATAAAGTACCTAGCAGATTTGGCTTATCTCCAGGGGCCCA	2542
Dp		1512	GAACTGGAAGGAGCAATAAAGTACCTAGCAGATTTGGCTTATCTCCAGGGGCCCA	1571
Qy		2543	CACCTCGTGTGAGCCATCATACAGAAAGTGAAGTGAATGATGACAGCAAGATGCC	2602
Dp		1572	CACCTCGTGTGAGCCATCATACAGAAAGTGAAGTGAATGATGACAGCAAGATGCC	1631
Qy		2603	ATCCCTCTTTTGTGTGGGGTATGGGTCGCAAAAGCCAAAGTGAACATTAATAATGT	2662
Dp		1632	ATCCCTCTTTTGTGTGGGGTATGGGTCGCAAAAGCCAAAGTGAACATTAATAATGT	1691
Qy		2663	ATTGAGC 2669	
Dp		1692	ATTGAGC 1698	

DR WPI: 2001-476164/51.
P-PSDB: AAM24252.

XX Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -

PS Claim 1: Page 664; 1275pp; English.

XX The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.

XX Sequence 580 BP; 146 A; 154 C; 116 G; 164 T; 0 other;

Query Match 10.9%; Score 291.6; DB 22; Length 580;
Best Local Similarity 97.1%; Pred. No. 3.4e-68;

Matches 297; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1509 ATGTTACTCAGAGAGGCAATGCTTTACTGAACTGGAAGAGCGGCTGACACACTGG 1568

Db 275 AGCTTACTCAGAGAGGCAATGCTTTACTGAACTGGAAGAGCGGCTGACACACTGG 334

QY 1569 TCCATATTTCTTTTCCCAAGATCTCGGGAGCTTCCGAGACATGCTCATCATCATCAT 1628

Db 335 TCCATATTTCTTTTCCCAAGATCTCGGGAGCTTCCGAGACATGCTCATCATCATCAT 394

QY 1629 CTAGGGATGCTCACCCTACTGCTGCTGCGCCACCTGAGCGGCTCCAGCCCTTCTGCT 1688

Db 395 CTAGGGATGCTCACCCTACTGCTGCTGCGCCACCTGAGCGGCTCCAGCCCTTCTGCT 454

QY 1689 GCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1748

Db 455 GCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 514

QY 1749 GGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1808

Db 515 GGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 574

QY 1809 CTACCT 1814

Db 575 CTACCT 580

RESULT 6

ABK51681 ID ABK51681 standard; DNA; 1920 BP.

AC ABK51681;

XX 30-JUL-2002 (first entry)

DE DNA encoding human ABCG5 protein.

XX Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
KW chromosome 2p21; ds.

XX Homo sapiens.

XX Key location/Qualifiers

FT CDS 1..1920

FT /product= "Human ABCG5 protein"

FT /transl_except= (pos: 4..9, aa: GDLSLIPGSGMGL)

XX /note= "This sequence contains 13 exons"

XX MO200227016-A2.

PD 04-APR-2002.

PF 25-SEP-2001; 2001WO-US29859.

XX 25-SEP-2000; 2000US-235268P.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (PATEL) PATEL S B.

XX (DEAN/) DEAN M.

PI Patel SB, Dean M;

XX WPI: 2002-416483/44.

DR P-PSDB: AAU98984.

XX Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
PT nucleic acid encoding the polypeptide, useful for treating
PT sitosterolemia, arteriosclerosis and heart diseases

PS Claim 38; Page 36-37; 66pp; English.

XX The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying
CC a compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present nucleic
CC acid sequence represents the human ABCG5 gene located on chromosome 2p21.
CC This sequence encodes the human ABCG5 protein of the invention.

XX Sequence 1920 BP; 440 A; 503 C; 486 G; 491 T; 0 other;

Query Match 7.6%; Score 203.6; DB 24; Length 1920;
Best Local Similarity 54.4%; Pred. No. 3.4e-44;

Matches 432; Conservative 0; Mismatches 359; Indels 3; Gaps 1;

QY 335 GCTGCCAGAAATTTCTGTGAGCTGGGCAATCCAGAACTTACCTTCAAGTGAAGTGGC 394

Db 143 GCCGGAGAGAGTGGAGCAGGAGATCTCAAAAGATGTCTCTTACGTGAGAGCGGGC 202

QY 395 AGATGCTGGCCATATATGAGGAGCTCAGGTTGTGGAGAGCCCTCTTCTAGATGATCA 454

Db 203 AGATGATGATGATCTTGTGGAAGATCCGAGCTCCGGGAAACACAGCTCTGAGCCATGT 262

QY 455 CTGGCCGAGGTCACGCGGAGGAGTCAAGTCAAGTCAGGCGCAGATCGATGATGAGGAGGCCA 514

Db 263 CCGGAGAGCTGGGGCGGCGGAGCTTCTGGGGAGGTGTATGTAAAGCGCGGGC 322

QY 515 GCTGCTCAGCTGTGTGAGGAAGTGTGCGCCACGTCGCGCAGACACACACAGCTGCTCC 574

Db 323 TCGCGCGGAGAGCAAGTTCACAGACTGCTTCTCTACGTCCTGAGAGAGACACCTGTGA 382

QY 575 CCAACTTGACTGTGCGAGAACCTTGGCTTTCATTTGCCAGATGCGGCTGCCAGAACT 634

Db 383 GCAGCTTCACCGTGGGAGACCTGTCATCACACCGGCTGTGGCCATCTCCGCGGGCA 442

QY 635 TCTGCCAGGCCACGCGGAGCAAAAGGTGTGAGAGCGATGCGGAGAGCTCGGCTTAAAGC 694

Db 443 ATCCCGGCTCTTCC---ACAAGAGGTGGAGGCGCTGATGAGAGCTGATGTAGCC 499

OY 695 AGTGGCTGACACCCGGTGGCAACATGATGCGGGGGTGTGCGGGGTGAGCCGA 754
 DB 500 ATGTGGAGACCGACTGATGTGGCAACTGACCTTGGGGGATTTCCAGGGGTAGCGGC 559
 OY 755 GGAAGTCATGATGGGGTGGAGCTCCTGTGTGAAACCCAGGAATCCTTATTTCTGACGAC 814
 DB 560 GCCGGGTCTCCATGCGACGCCAGCTGTCCAGAGATCCAGTCAATGCTGTTTATGTAGC 619
 OY 815 CCACCTCTGGGCTGACAGCTTTCACAGGCCCAACACCTGTTGTAAGACCTTGTCCAGCTGG 874
 DB 620 CAACACAGGCTGTGACTGCTGATGCTGTAATGATGCTGCTCTGCTGTGAACTGG 679
 OY 875 CCAAGGCAACCGGCTGTGCTCAATCTCCCTCCACAGCCCTGCTGTGACATCTTCAGGC 934
 DB 680 CTCGACGAAACCGAATGTGTGTTCTCAACATTCACAGCCCGCTTCTGAGCTTTTTCAGC 739
 OY 935 TGTGTGATGTCGTCCTCTGATGACGCTGTGCGACCCCGCACTTACATTAAGGGCGCCGAC 994
 DB 740 TCTTTGACAAATTTGCAATCTGAGCTTCCGAGAGCTGATTTTCTGTGACGCGCAGCGG 799
 OY 995 ACATGCTCAGTATTTTCACAGCCATGCGGCTACCCCTGCTCCGTACAGCAATCTGCTG 1054
 DB 800 AAATGCTGATTTTTCATGACTGCGGTTACCTTGTCTGAAACATTCAAACCTTTTG 859
 OY 1055 ACTTCTATGTGACCTGACAGCAATGACAGGCGCAGCAGAGACAGCAATTTGCCACCA 1114
 DB 860 ACTTCTATGTGACCTGACAGCAATGACAGGCTGATACCCAAACAGAGAAATGAAACTG 919
 OY 1115 GCGGAGAGCTCAG 1128
 DB 920 CCAAGAGATCCAG 933
 RESULT 7
 AAD22009
 ID AAD22009 standard; DNA: 2340 BP.
 AC AAD22009;
 DT 12-FEB-2002 (first entry)
 DE Human sitosterolemia susceptibility gene (SSG).
 XX Human: sitosterolemia susceptibility gene; SSG: atherosclerosis;
 KM sterol-related disorder; hyperlipidemia; hypercholesterolemia; therapy;
 KM gall stone; coronary heart disease; cardiovascular disease; arthritis;
 XX xanthoma; hemolytic anaemia; transgenic animal; chromosome 2p21; ds.
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH 107.2062
 FT /tag- a
 FT /product- "Human SSG protein"
 XX
 XX MO200179272-A2.
 PD 25-OCT-2001.
 XX
 XX 18-APR-2001; 2001MO-0512758.
 XX
 XX 18-APR-2000; 2000US-198465P.
 PR 15-MAY-2000; 2000US-204234P.
 XX
 XX (TULA-) TUTARIK INC.
 XX
 XX Tian H, Schultz J, Shan B;
 XX
 XX MPI: 2002-017598/02.
 DR P-PSDB: AAE13290.
 XX
 XX Novel sitosterolemia susceptibility gene polypeptide and
 PT polynucleotide, useful for screening a compound that increases the

PT level of expression or activity of SSG polypeptide for treating
 PT sterol-related disorder
 XX
 XX Claim 8; Fig 8; 105pp; English.
 PS
 CC The invention relates to an isolated sitosterolemia susceptibility gene
 CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
 CC binding cassette (ABC) family cholesterol transporter. SSG is useful
 CC for identifying a compound useful in the treatment or prevention of a
 CC sterol-related disorder, including sitosterolemia, hyperlipidemia,
 CC hypercholesterolemia, gall stones, HDL deficiency, atherosclerosis or
 CC nutritional deficiencies. SSG is also useful for treating cholesterol-
 CC associated diseases or conditions including coronary heart disease and
 CC other cardiovascular diseases, and sitosterolemia-associated condition
 CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
 CC expression cassette is useful in the production of transgenic non-human
 CC animals. SSG genes and their homologues are useful as tools for a number
 CC of applications including diagnosing sitosterolemia and other
 CC cardiovascular disorders, for forensics and paternity determinations,
 CC and for treating any of a large number of SSG associated diseases. The
 CC present sequence is human SSG DNA. Human SSG is located on chromosome
 CC 2p21.
 XX
 XX Sequence 2340 BP; 541 A; 601 C; 598 G; 600 T; 0 other;
 Query Match 7.6%; Score 203.6; DB 24; Length 2340;
 Best Local Similarity 54.4%; Pred. No. 3.8e-44;
 Matches 432; Conservative 0; Mismatches 359; Indels 3; Gaps 1;
 OY 335 GCTGCGAATTTCTGTGATGGGCAATCCAGAACCTTAAGTGAAGTGGC 394.
 DB 285 GCCGCGACAGTACGACAGCAGATCTCTCAAGATGTCCTGTAGCGAGCGGC 344
 OY 395 AGATGCTGCCATCATAGGAGGAGCTCAGTGTGTGGAGAGCTCTCTAGATGTATCA 454
 DB 345 AGATCATGTGATCTCTAGGAAGCTCAGAGCTCCGGGAAACACAGCTCTGAGCCCATGT 404
 OY 455 CTGGCCAGAGTACAGCGCGCAATCAATGAGCCAGATCTGATCAATGGCAGCCCA 514
 DB 405 CCGGAGAGCTGGGGCGCGCGGAGCTTCTGTGGGGAGTGTATGTAAGCGCGGCGC 464
 OY 515 GCTGCGCTCAGCTGTGAGAGAGTGTGGCCAGCGCCCGCAACACAGCTGCTCC 574
 DB 465 TCGCCCGGAGCACTTCCAGACTGCTTCTCTACGCTCCAGAGGAGAACCTGTGA 524
 OY 575 CCACTTGTAGTGTGCGAGAGACCTTGTGCTTCAATGGCCAGATGCGGCTGCCAGACT 634
 DB 525 GCAAGCTCAGCGTGGGAGAGCGTGCACATACCGGCTGCGCATCCGCGCGCA 584
 OY 635 TCTCCAGGCCAGCGTGCACAAAGGTGAGAGAGTATCGCGGAGCTCGGCTTATGAGC 694
 DB 585 ATCCCGGCTCTTCC--AGAGAGAGGTGAGGCGCTCATGCGAGACTGATGCTGAGCC 641
 OY 695 AGTGGCTGACACCCCGTGGCAACATGATGCTGCGGGGCTGTGCGGGGTGAGCGCA 754
 DB 642 ATGTGCGACCGAGCTGATGATGCAATGAGCTTGGGGGCAATTTCCACGCTGAGGCGC 701
 OY 755 GGAAGTCATGATGGGGTGGAGCTCCTGTGGAACCCAGGAATCCTTATTTCTGAGGAC 814
 DB 702 GCCGGGTCTCCATGCGACGCCAGCTGTCCAGATCTTAAGTCAATGCTTTGTATGAGC 761
 OY 815 CCACCTCTGGGCTGACAGCTTTCACAGGCCCAACACCTGTGAGAGACTTGTCCAGCTGG 874
 DB 762 CAACACAGGCTGTGACTGCTGATGCTGTAATGATGCTGCTCTGCTGTGAACTGG 821
 OY 875 CCAAGGCAACCGGCTGTGCTCAATCTCCCTCCACAGCCCTGCTGTGACATTTAGGC 934
 DB 822 CTCGACGAAACCGAATTTGTGTCTCACATTCACAGCCCGCTTCTGAGCTTTTTCAGC 881
 OY 935 TGTGTGATGTCGTCCTCTGATGACGCTGTGCGACCCCGCACTTACATTAAGGGCGCGCAGC 994
 DB 882 TCTTTGACAAATTTGCCATCTGAGCTTCCGAGAGCTGATTTTCTGTGCGACCGCAGCGG 941

OY 995 ACATGTCAGATTTCACAGCCATCGGCTACCCCTGCTCTCTACAGCAATTCCTCTG 1054
 DB 942 AATGCTGATTTCTTAATGACTGCGCTTACCTGCTGAAACATTCAAACCTTTTG 1001
 OY 1055 ACTTCATGTGACCTGACAGCATTCACAGGCGACAGAGAGCAATTCGCGACCA 1114
 DB 1002 ACTTCATGTGACCTGACAGCATTCACAGGCGAAAGCAAGCAAGGAAATAGAACT 1061
 OY 1115 GGGAGAGGCTCAG 1128
 DB 1062 CCAAGAGAGTCCAG 1075
 RESULT 8
 AAD48882
 ID AAD48882 standard; DNA; 2340 BP.
 XX
 AC AAD48882;
 XX
 DT 24-MAR-2003 (first entry)
 XX
 DE Human ABCG5 DNA.
 XX
 KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;
 KM sitosterolemia; hyperlipidemia; hypercholesterolemia; gall stone;
 KM HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;
 KM human; ATP-binding cassette; sitosterolemia susceptibility gene; SSG;
 KM ABCG5; gene; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 107..2062
 FT /tag= a
 FT /product= "ABCG5 protein"
 FT
 FT
 PN W0200281691-A2.
 PD 17-OCT-2002.
 PE 20-NOV-2001; 2001MO-US43823.
 PR 20-NOV-2000; 2000US-252235P.
 PR 28-NOV-2000; 2000US-253645P.
 XX
 PA (TULAR) TULARIK INC.
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Hobbs HH, Shan B, Barnes R, Tian H;
 DR MPI: 2003-058548/05.
 DR P-PSDB: ABE31704.
 XX
 PT New ABCG8 polypeptides and nucleic acids, useful for treating
 PT sterol-related disorders e.g. sitosterolemia, hypercholesterolemia,
 PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
 PT nutritional deficiencies.
 XX
 PS Claim 11; Page 77; 94pp; English.
 XX
 CC The invention relates to ATP-binding cassette (ABC) family cholesterol
 CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
 CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also
 CC known as sitosterolemia susceptibility gene (SSG). Sequences of the
 CC invention are useful for treating or preventing sterol-related disorders
 CC such as sitosterolemia, hyperlipidemia, hypercholesterolemia, gall
 CC stones, HDL deficiency, atherosclerosis and nutritional deficiencies.
 CC They are also useful in gene therapy. The present sequence is human
 CC ABCG5 DNA.
 XX
 SO Sequence 2340 BP; 541 A; 601 C; 598 G; 600 T; 0 other;
 Query Match 7.6%; Score 203.6; DB 25; Length 2340;

Best Local Similarity 54.4%; Pred. No. 3.8e-44;
 Matches 432; Conservative 0; Mismatches 359; Indels 3; Gaps 1;
 OY 335 GCTGCCAGAAATTCCTTGAGCTGGGCAATCCAGAAACCTCAACTCAAAAGTGAAGTGGCC 394
 DB 285 GCCGGGACAGATGGAGCCAGCAATCTTCAAAAGTGTCTCTGTAGCTGAGAGCGCGC 344
 OY 395 AGATGCTGGCCATCATATGAGAGCTCAGTGTGTGGAGAGGCTCTCTGATGATGATCA 454
 DB 345 AGATCATGTGCATCTCTAGAGAGCTCAGAGCTCCGGGAAACACAGCTGTGTGAGCCATGT 404
 OY 455 CTGGCCGAGGTCAACGGCGGCAAGATCAAGTCAAGCCAGATCTGATCAATGGGAGCCCA 514
 DB 405 CCGGAGAGGCTGGGGGCGCGGAGACCTTCTGGGGAGGTGTATGTGAAGCGCGCGCGC 464
 OY 515 GCTGCGCTCAGCTGTGAGAGAAATGTGTGGCCCACTGGCCAGCAACACCTGTCTCC 574
 DB 465 TGGCGCGGAGCAATTCACAGACTCTTCTTCACTCAGCTCCAGAGAGCAACCTCTGTGA 524
 OY 575 CCAACTTGACTGTGACAGAGACCTTGACCTTCAATGCCCCAGATGCGGCTGCCAGAACT 634
 DB 525 GCAAGCTCAACCGTGGCGGAGACCTGTGACATACACCGCTGCTGCGCATCCGCCGGCA 584
 OY 635 TCTCCAGAGCCCAAGCTGACAAAGGTTGAGAGACGTGATCGGAGCTGCGCTTATAGCC 694
 DB 585 ATCCCGGCTCTCTCC---AGAAAGAGTGGAGGCCCTCATGACAGAGCTGATGTAGCC 641
 OY 695 AGTGGCTGACACCCCGGTGGGCAACATGATGACTGCGGGGTTGTGGGGGTTGAGCGCA 754
 DB 642 ATGTGGCAAGCCAGATGATTTGGCAACTACAGCTTGGGGGCAATTTCCAGCGGTGAGCGCG 701
 OY 755 GGAGAGTCAGACATTTGGGTTGACAGCTCTGTGAGAACCCAGAAATCTTATTCGAGAGAC 814
 DB 702 GCCGGTCTTCATCCGACGCCCACTCTCCAGAGATCTTAAGTCAATGCTGTGATGATAGC 761
 OY 815 CCACCTCTGGGCTGACAGCTTCAACAGCCCAACCTGTGAGAACCTTGTCCAGGCTG 874
 DB 762 CAACCAAGAGCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 821
 OY 875 CCAAGGCAACCGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 934
 DB 822 CTGCGAGAGCAACGAAATGTGTTCTACCACTTACAGCCGCTTCTGAGCTTTTTCACG 881
 OY 935 TGTGTGATCTGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 994
 DB 882 TCTTTGACAAATTTGCAATCTGAGCTTGGAGAGTGAATTTCTGTGACAGCCAGCGG 941
 OY 995 ACATGCTCAGATTTTCACAGCCATCGCTACCCCTGTCTGCTACAGCAATTCCTGCTG 1054
 DB 942 AATGCTGATTTCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1001
 OY 1055 ACTTCATGTGACCTGACAGCATTCACAGGCGACAGAGAGCAATTCGCGACCA 1114
 DB 1002 ACTTCATGTGACCTGACAGCATTCACAGGCGAAAGCAAGCAAGGAAATAGAACT 1061
 OY 1115 GGGAGAGGCTCAG 1128
 DB 1062 CCAAGAGAGTCCAG 1075
 RESULT 9
 ABE51682
 ID ABE51682 standard; cDNA; 2516 BP.
 XX
 AC ABE51682;
 XX
 DT 30-JUL-2002 (first entry)
 XX
 DE Human ABCG5 cDNA sequence.
 XX
 KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KM atherosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KM chromosome 2p21; ss.

XX Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 XX Example 3; Page 45-46; 66pp; English.

CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present nucleic
 CC acid sequence encodes the rat ABCG5 protein of the invention.

XX Sequence 2035 BP; 481 A; 533 C; 537 G; 484 T; 0 other;

Query Match 7.3%; Score 194.4; DB 24; Length 2035;

Best Local Similarity 54.0%; Pred. No. 1, 1e-41;

Matches 421; Conservative 0; Mismatches 356; Indels 3; Gaps 1;

QY 360 CATCCAGAACCTTAAGCTTCAAGTGAAGAGGAGATGCGCCATCATAGGAGCTC 419
 DB 214 CCTCAAGATGCTCTTGTATCATGAGAGTGGCAGACCATGTGATAGTACGTC 273
 QY 420 AGGTGTGGGAGAGCTCTCTTGTATGATGATGATGATGATGATGATGATGAT 479
 DB 274 AGGCTCAGGGAAGAACGAGCTGCTGAGCCATCTGTGGAGAGTGGCGGACAG 333
 QY 480 CAACTCAGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 539
 DB 334 CTTGGAAGGGAAGTCTTGTATGATGATGATGATGATGATGATGATGATGATG 393
 QY 540 TGTGCGCAAGCTGCGCAGACACACACACACACACACACACACACACACAC 599
 DB 394 GGTCTCTTACCTCTCTGAGAGAGATGCTTCTGAGCAGACCTCAGCGGGAAC 453
 QY 600 GGCCTTCATTTGCCAGATGCGGCTCCCAAGACCTTCTCCAGGCGCCAGCTG 659
 DB 454 GAGATACAGGC--GATGCTGGCTCTCCGACAGCTCCGGGAGCTTCTACGAC 510
 QY 660 GGTGAGAGAGCTGATCGGAGAGCTGCGCTTAGGAGTGGCTGACACCCGCG 719
 DB 511 GGTAGAGAGAGCTCTGAGAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 570
 QY 720 CATGATGCTGCGGGGCTTGTGCGGGGATGAGCGCAGAGATGATGAGCTG 779
 DB 571 CTATTAATTTTGGGGGATTTCCAGTGGCGAGCGGCGGAGTGTCAATCGAC 630
 QY 780 CTTTGGGAACCCAGGAATCTTATTTCTGAGAGACCCACCTCTGCGGCTGAC 839
 DB 631 CTTTGGAGAGCCCAAGTCAATGATGATGATGATGATGATGATGATGATGATG 690
 QY 840 AGCCACAAACCTGTGTAAGACCTTGTCCAGAGCTGCGCCCAAGAGCAGCG 899
 DB 691 TGCAATGATATTCCTCTCTCTTGTGATGATGATGATGATGATGATGATGAT 750
 QY 900 CTCCTCTCAGCAGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 959
 DB 751 CACATCCACAGCAGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 810

QY 960 GTCTGCACCCCATCTACTTAAAGGCGGCCAGACATGCTCAGTATTTACAGCAT 1019
 DB 811 TTACGAGAGATGCTGTCTGTGTCAGCAGCAGAGAGATGCTGCTTCTCATAC 870
 QY 1020 CGGCTACCCCTGTCTCTGCTTACAGCAATCTGCTACTTATGAGACCTGAC 1079
 DB 871 TGGTTACCCCTGTCTGTAACATCTTGTATTTCTACATGACTTGCATTCG 930
 QY 1080 TGACAGCGCAGCAGAGAGATGAGTGGCCACAGGAGAAAGCTTACGCTGAC 1139
 DB 931 GGACACCAAGCAGAGAGAGATGAGTGGCCACAGGAGAAAGCTTACGCTGAC 990

RESULT 11

ABK51684

ID ABK51684 standard; DNA; 1915 BP.

XX ABK51684;

XX 30-JUL-2002 (first entry)

XX DNA encoding mouse ABCG5 protein.

XX Mouse; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW ds.

XX Mus sp.

XX Key Location/Qualifiers
 FH 1..1915
 FT CDS

XX /tag- a

XX /product- "Mouse ABCG5 protein."
 FT /transl_except- (pos: 1912..1915, aa: LGIVIFRVRLISR)
 FT /note- "This sequence lacks a stop codon."

XX W020227016-A2.

XX 04-APR-2002.

XX 25-SEP-2001; 2001WO-US29859.

XX 25-SEP-2000; 2000US-235268P.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX (PATEL) PATEL S B.

XX (DEAN/) DEAN M.

XX Patel SB, Dean M;

XX WPI: 2002-416483/44.

XX P-PSDB: AAN96985.

XX Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases

XX Example 3; Page 42-43; 66pp; English.

CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The

CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present nucleic
 CC acid sequence encodes the mouse ABCG5 protein of the invention.

SO Sequence 1915 BP; 453 A; 502 C; 484 G; 476 T; 0 other;

Query Match 7.2%; Score 193.4; DB 24; Length 1915;
 Best Local Similarity 53.4%; Pred. No. 1.9e-41;
 Matches 429; Conservative 0; Mismatches 371; Indels 3; Gaps 1;

OY 335 GCTGCCAATTTCTGTGAGCTGGGCAATCCAGAACCTTAAGCTTCAAGTGAAGAGTGGGC 394
 DB 182 GCCAGCAGAGTGGGACAGGCAAAATCTCAAGATGCTCTCTGTACATCGAGAGTGGCC 241
 OY 395 AGATGCTGGCCATCATAGAGCTGAGGTTGGGAGAGCCTCTCTGTAGATGATGATCA 454
 DB 242 AGATTATGTGATCTTAAAGAGCTGAGCTGAGGAGAGACCAAGCTGCTGAGAGCCATCT 301
 OY 455 CTGGCCGAGTCAAGGCGGCAAGATCAAGTCAAGCTGATCTGATCAATGAGGAGCCCA 514
 DB 302 CCGGAGAGCTGGGCGGCACTGGGACCTGGAGAGGAGAGTGTCTTGTGAATGGCTCCGAGC 361
 OY 515 GCTGCGCTCAGCTGTGAGAGATGTGTGGCCCAAGCTGCGCAGCAGCAGCAGCTGCTCC 574
 DB 362 TCGCGAGGAGCAGTCCAGAGCTGCTCTCTAGCTCAGAGCAGCAGCTTTTCTGA 421
 OY 575 CCAACTTACTGTGAGAGAGACTTGGCCCTTATGGCCCAATGAGGCTGGCCAGAACT 634
 DB 422 GCAGCTCAGTGTGGCGAGAGCTTGGATACAGC--GATGTGGCCCTTCTCCGCA 478
 OY 635 TCTCCAGAGCCCAAGCTGACAAAGAGGTGAGAGAGTGTGCGGAGTGGGCTTAAAGC 694
 DB 479 GCTCCGCGAGCTTCTACAAAGAGAGTGAAGAGTATACAGAGCTGAGCCTGAGCC 538
 OY 695 AGTGGCTGAGACCCGCTGGCAATGATGACGTGCGGGGTTGTGCGGGGTGAGCGCA 754
 DB 539 ACGTGGCGGCAAAATGATGTGGCAGTAAATTTTGGGGAAATTTCCAGTGGCAGCGGC 598
 OY 755 GGAGTCAAGATGGGGTGGAGCTGCTGAGAACCCAGAAATCTTATTTTCCAGCAAC 814
 DB 599 GCCAGTTTCCATGCAAGCCCAACTCTTCAGAGACCCCAAGGTATGATGATGATGAGC 658
 OY 815 CCACCTTGGGCTGAGAGCTTTCACAGCCCAACACTGTGTGAAGACCTTGTCCAGGGCG 874
 DB 659 CAACACAGAGCTGAGTGCATGATGATCAATTAATTTCTCTGTGCTGAGCTGAGC 718
 OY 875 CCAAGGCAACCGGCTGTGCTATCTCCACCAAGCTGCTGTGATCTTCAAGC 934
 DB 719 CTCGAGGAGCCGATGTGATTTCTCAACATCCAGACGCTGCTGAGCTTCTTCAAC 778
 OY 935 TGTGATGCTGCTCTGATGCTGAGCTGAGACCCCACTTACTTATGAGGGGCGCCAGC 994
 DB 779 ACTTGAGCAAAATTTCCATCTCTGACTTACGAGAGTGTGTGTTGAGCCCAAGG 838
 OY 995 ACATGCTCAGTATTTTCACAGCCATCGGTAACCCCTCTCTGCTTACAGCAATCTCTGT 1054
 DB 839 AGATGCTTGGCTTCTCAATAAATCTGTGTTACCCCTCTCTGTAACATTTCCATCTCTTTG 898
 OY 1055 ACTTCTATGTGAGCTTCAAGCAATTCAGAGGCGCAGAGAGAGAGAGAGAAATGGCCACA 1114
 DB 899 ATTTTACATGAGCTTCAATCAATCAAGTGAACACCAAGAGAGAGAGAGAGAGAAATGAAACGT 958
 OY 1115 GGGAGAGGCTCAGTCACTGCA 1137
 DB 959 ACAGAGAGTACAGATCTGGA 981

RESULT 12

AAD48880
 ID AAD48880 standard; DNA; 1959 BP.
 XX
 AC AAD48880;
 XX
 XX 24-MAR-2003 (first entry)
 DT
 XX
 DE Mouse ABCG5 DNA.
 XX
 XX ABC family cholesterol transporter; ABCG8; sterol-related disorder;
 KW sitosterolemia; hyperlipidemia; hypercholesterolemia; gall stone;
 KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;
 KW mouse; ATP-binding cassette; sitosterolemia susceptibility gene; SSG;
 KW ABCG5; gene; ds.
 XX
 XX Mus sp.
 OS
 XX
 XX Key Location/Qualifiers
 FH 1.1591
 FT CDS /*tag=a
 FT /product="mABCG5 protein"
 PN
 PD
 PF 17-OCT-2002.
 PF 20-NOV-2001; 2001WO-US43823.
 PR 20-NOV-2000; 2000US-252235P.
 PR 28-NOV-2000; 2000US-253645P.
 XX
 XX (TULAR) TULARIK INC.
 PA (TEXA) UNIV TEXAS SYSTEM.
 PI
 PI Hobbs HH, Shan B, Barnes R, Tlan H;
 DR WPI: 2003-058548/05.
 DR P-PSDB: AAE31702.
 XX
 XX New ABCG8 polypeptides and nucleic acids, useful for treating
 PT sterol-related disorders e.g. sitosterolemia, hypercholesterolemia,
 PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
 PT nutritional deficiencies -
 XX
 XX Claim 11; Page 73; 94pp; English.
 PS
 PS
 PS
 CC The invention relates to ATP-binding cassette (ABC) family cholesterol
 CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
 CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also
 CC known as sitosterolemia susceptibility gene (SSG). Sequences of the
 CC invention are useful for treating or preventing sterol-related disorders
 CC such as sitosterolemia, hyperlipidemia, hypercholesterolemia, gall
 CC stones, HDL deficiency, atherosclerosis and nutritional deficiencies.
 CC They are also useful in gene therapy. The present sequence is mouse
 CC ABCG5 DNA.
 XX
 SO Sequence 1959 BP; 468 A; 506 C; 495 G; 490 T; 0 other;

Query Match 7.2%; Score 193.4; DB 25; Length 1959;
 Best Local Similarity 53.4%; Pred. No. 2e-41;
 Matches 429; Conservative 0; Mismatches 371; Indels 3; Gaps 1;

OY 335 GCTGCCAATTTCTGTGAGCTGGGCAATCCAGAACCTTAAGCTTCAAGTGAAGTGGGC 394
 DB 182 GCCAGCAGAGTGGGACAGGCAAAATCTCAAGATGCTCTCTGTACATCGAGAGTGGCC 241
 OY 395 AGATGCTGGCCATCATAGAGCTGAGGTTGGGAGAGCCTCTCTGTAGATGATGATCA 454
 DB 242 AGATTATGTGATCTTAAAGAGCTGAGCTGAGGAGAGACCAAGCTGCTGAGAGCCATCT 301
 OY 455 CTGGCCGAGTCAAGGCGGCAAGATCAAGTCAAGCTGATCTGATCAATGAGGAGCCCA 514
 DB 302 CCGGAGAGCTGGGCGGCACTGGGACCTGGAGAGGAGTGTCTTGTGAATGGCTCCGAGC 361

```

OY 515 GCTGCGCTCAGCTGTGAGAGAGTGTGGCCGACGCGCCAGACACACACAGCTGCTCC 574
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 TGGCGAGGAGACACAGTTCACAGACTGCTTCTCAGCTGCTGACAGAGAGAGCTTTTCTGA 421
OY 575 CCAACTGACTGTGCGAGAGACCTTGCCCTTCAATGCGCCAGATGCGGCTCCAGAACCT 634
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 422 GCAGCTCAGCTGTGCGGAGAGCTTGCGATACACAGC---GATGCTGCGCTTGTCCGCA 478
OY 635 TCTCCAGGCGCCAGCTGACAAAAGGCTGAGAGAGCTGATGCGGAGCTGCGCTTAGGC 694
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479 GCTCCGCGGACTTCTACACACAGAGAGAGAGCTGATGACAGAGCTGAGCTGAGCC 538
OY 695 AGTGGCTGACACCGCGGTGGGACATGACCTGCGGGGCTGTGCGGGGCTGAGCGCA 754
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 539 ACCTGGCGGACCAATGATTTGGCAGCTATATATTTTGGGGAATTTCCAGTGGGAGCGC 598
OY 755 GGGAGTCAAGATTGGGGTGTGAGCTCTGTGAAACCCAGAGATCTTATTTCTGAGGAC 814
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 599 GCCGAGTTCATCCGACGCCAACCTCTTCAGAGACCCCAAGCTCATGATGATGATGAGC 658
OY 815 CCACCTCTGGCTCGACAGCTTCACAGCCCAACCTGTGAGACCTTGTCCAGGCTGG 874
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 659 CAACCAAGAGAGCTGAGCTGATGACGAAATCAATTTGCTTCTTGGCTGAGCTGG 718
OY 875 CCAAGGCAACCGGCGTGCATCTCCCTCCACAGAGCTGCTGTGACATCTTACAGC 934
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 719 CTCGAGGAGACCGAATGTGATGTACCATCCACAGCTGCTGTGAGCTCTTCCACAC 778
OY 935 TGTGTGATCTGCTCTCTGATGACGTCGTGCAACCCCACTACTTATGAGGGGGGCCAC 994
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 779 ACTTGACAAATTCGATCTGATCTGACGAGAGTGTGTGTGTGACCCCAAGAG 838
OY 995 ACATGCTCAGATTTTACAGAGCCATCGCTACCCCTGCTGCTGCTACAGCAATCTGCTG 1054
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 839 AGATCTCTGCTCTTCTCATATACGTGTGTTACCCCTGCTGACATTTCCATCTTGG 898
OY 1055 ACTTGTGAGTGAAGCTGACCCAGCATTTGACAGCCGACGAGAGAGCAATGGCCACCA 1114
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 899 ATTTTACATGAGCTGACATGACAGCCCAAGCAGAGAGCGGGAATAGAAACGT 958
OY 1115 GGGAGAGGCTCAGTCACTCCCA 1137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 959 ACAAGCAGTACAGATGCTGGAA 981

```

RESULT 13

AAD22008
ID AAD22008 standard; DNA; 2258 BP.

AC AAD22008;

DT 12-FEB-2002 (first entry)

DE Mouse sitosterolemia susceptibility gene (SSG).

XX Mouse: sitosterolemia susceptibility gene; SSG; atherosclerosis;

XX sterol-related disorder; hyperlipidaemia; hypercholesterolemia; therapy;

XX gall stone; coronary heart disease; cardiovascular disease; arthritis;

XX xanthoma; haemolytic anaemia; transgenic animal; chromosome 17; ds.

OS Mus sp.

FT Key Location/Qualifiers

FT CDS 47..2005

PF 18-APR-2001; 2001WO-US212758.

```

XX 18-APR-2001; 2000US-198465P.
PR 15-MAY-2000; 2000US-204234P.
XX
PA (TULA-) TULARIK INC.
XX Tian H, Schultz J, Shan B;
PI MPI; 2002-017598/02.
DR P-PSDB; AAE13289.
XX
PT Novel sitosterolemia susceptibility gene polypeptide and
PT polynucleotide, useful for screening a compound that increases the
PT level of expression or activity of SSG polypeptide for treating
PT sterol-related disorder
XX
PS Claim 8; Fig 7; 105bp; English.
XX
CC The invention relates to an isolated sitosterolemia susceptibility gene
CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
CC binding cassette (ABC) family cholesterol transporter. SSG is useful
CC for identifying a compound useful in the treatment or prevention of a
CC sterol-related disorder, including sitosterolemia, hyperlipidaemia,
CC hypercholesterolemia, gall stones, HDL deficiency, atherosclerosis or
CC nutritional deficiencies. SSG is also useful for treating cholesterol-
CC associated diseases or conditions including coronary heart disease and
CC other cardiovascular diseases, and sitosterolemia-associated condition
CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
CC expression cassette is useful in the production of transgenic non-human
CC animals. SSG genes and their homologues are useful as tools for a number
CC of applications including diagnosing sitosterolemia and other
CC cardiovascular disorders, for forensics and paternity determinations,
CC and for treating any of a large number of SSG associated diseases. The
CC present sequence is mouse SSG DNA. Mouse SSG is located on chromosome 17.
XX
SQ Sequence 2258 BP; 549 A; 579 C; 567 G; 563 T; 0 other;

```

Query Match 7.2% Score 193.4; DB 24; Length 2258;
Best Local Similarity 53.4%; Pred. No. 21e-41;
Matches 429; Conservative 0; Mismatches 371; Indels 3; Gaps 1;

```

OY 335 GCTGCCAGAAATCTTGTGAGTGGAGTGGACATCCAGAACCTTAAGCTCAAAAGTGAAGTGGG 394
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 228 GCCAGCAGAGAGTGGGACAGCGCAAAATCTCAAAAGATGTCTCTTGTACATCCAGAGTGGCC 287
OY 395 AGATGCTGGCCATCATATAGGAGCTGAGCTTGTGGAGAGCCTCTTGCTAATGTGATCA 454
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 288 AGATTATGTGATCTTATAGCAGCTCAGGCTCAGGGAGAACACACAGCTGCTGAGACGCATCT 347
OY 455 CTGGCGAGGTCACGCGGCGCAAGATCAAGTACAGGCCAGATGTGATCAATGAGGACCCA 514
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 348 CCGGAGAGCTCGGCGCCTGAGACCTGGAAGGAGAGTGTGTGTGATGCTGCGAGC 407
OY 515 GCTGCGCTCAGCTGTGAGAGAGTGTGAGCCACAGTGGCGACACACACAGAGTGTCC 574
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 408 TGGCAGAGGACACAGTTCACAGACTGCTTCTTACAGTCTCCAGACAGCACTTTTCTGA 467
OY 575 CCAACTGACTGTGCGAGAGACCTTGCCCTTCAATGCCCAGATGCGGCTGCCAGAACT 634
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 468 GCAGCTTCACTGTGCGGAGACGTGTGGATACACAGC---GATGCTGGCCCTTGCGGCA 524
OY 635 TCTCCAGGCGCCAGCTGACAAAAGGGTGGAGAGAGCTGATGCGGAGCTGCGGCTTAGGC 694
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 525 GCTCGGCGGACTTATACAAAGAGATGAGGAGATGACAGAGCTGAGCTGAGCC 584
OY 695 AGTGGCTGACACCGCGGTGGGCAACATGTACGTGCGGGGCTGTGCGGGGCTGAGCGCA 754
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 585 ACGTGGCGGACCAATGATTTGAGCAGCTATATATTTTGGGGAATTTCCAGTGGGAGCGC 644
OY 755 GGAGAGTCAAGATTTGGGCTGAGCTCTGTGGAACCCAGCAATCTTATTTCTGAGCAAC 814
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 645 GCCGAGTTTCCATGCGAGCCCACTCTTCAAGAGACCCCAAGGTCATGATGATGAGC 704

```


OY 815 CCACCTGGGGCTGCAGACCTTCACAGCCCAACCTGGTGAAGACCTTGTCCAGCTGG 874
 DB 705 CAACACAGAGACTGCATGCATGATGCAGAAATTCATATTCCTTCTTGGCTGAGCTGG 764
 OY 875 CCAAGGCAACCGGCTGCTGCATCTCCCTCCACAGCCCTGCTGATCATCTTCCAGGC 934
 DB 765 CTCGAGGAGCCGAAATTCATGTTGATCCATCCACAGCCCTGCTGAGCTTCCCAAC 824
 OY 935 TGTTCATCTGCTCTCCGATGAGCTGAGACCCCATCTGATAGGCGCGCCAGC 994
 DB 825 ACTTCGACAAATTCCTCCATCTGATTCAGGAGAGTGTCTGTGCGACCCAGAGG 884
 OY 995 ACATGCTCAGATATTCACAGCCATCGCTACCCCTGCTGCTACAGCAATCTGCTG 1054
 DB 885 AGATGCTGCTGCTTCTCAATACGTGTGTTACCCCTGCTGCAATTCCTCATCTGCTTG 944
 OY 1055 ACTTCATGATGAGCTGACACGATTCAGAGCGGAGAGAGAGAGAAATTCGCCACCA 1114
 DB 945 ATTTTACATGAGCTGATGATCAGTGAGACCCCAAGAGAGAGCGGGAATAGAAAGCT 1004
 OY 1115 GGGAGAGGCTCAGTCACTGCCA 1137
 DB 1005 ACAAGCGAGTACAGATGCTGGA 1027
 RESULT 14
 ABR51685
 ID ABR51685 standard; cDNA: 2354 BP.
 AC ABR51685;
 XX 30-JUL-2002 (first entry)
 DE Mouse ABCG5 cDNA: sequence.
 KW Mouse; ABCG5, ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW ss.
 OS Mus sp.
 XX W0200227016-A2.
 PN 04-APR-2002.
 XX 25-SEP-2001; 2001MO-US29859.
 PE 25-SEP-2000; 2000US-235268P.
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 PI Patel SB, Dean M;
 XX WPI: 2002-416483/44.
 DR Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 PS Example 3; Page 45; 66pp; English.
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or

CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present nucleic
 CC acid sequence represents the cDNA sequence of the mouse ABCG5 gene of
 CC the invention.
 XX
 SO Sequence 2354 BP; 573 A; 604 C; 594 G; 583 T; 0 other;
 Query Match 7.2%; Score 193.4; DB 24; Length 2354;
 Best Local Similarity 53.4%; Pred. No. 2.2e-41;
 Matches 429; Conservative 0; Mismatches 371; Indels 3; Gaps 1;
 OY 335 GCTGCCAATTTCTGTAGCTGGGATCCAGAACTTAAGTTCAATGAGAACTGGGC 394
 DB 320 GCCAGCAGAACTGGGAGCAGCAAAATCCCAAGATGCTCTTGTACATCGAGAGTGGCC 379
 OY 395 AGATGCTGGCATCATAGGAGCTCAGGTTGGGAGAGCCGCTGTAGATGATCA 454
 DB 380 AGATTATGATCTTATAGGACGCTCAGGCTCAGGAGAACACAGCTGCTGAGCGCATCT 439
 OY 455 CTGGCCGAGGTGACGGCGCAATCAAGTCAAGGCGCATGCTGATCAATGGGAGCCCA 514
 DB 440 CCGGAGAGCTGCGGGCAGCTGGAGCCCTGGAAGGGAGGTGTTGTATGATGCTGGAGC 499
 OY 515 GCTGGCCCTCAGCTGTGAGGAAGTGTGCGCCACGTGCGCAGACACCAAGCTGCTCC 574
 DB 500 TCGCAGAGGACCAATTCACAGCTGCTCTCTCTCCTCAGACAGGACGTTTTCGA 559
 OY 575 CCAACTGACTGTGAGAGAGCTTGGCTTCAATGGCCAGATCGGCTGCCAGAACT 634
 DB 560 GCACCTCAGCTGTGAGAGAGCTTGGCTTCAATGGCCAGATCGGCTGCCAGAACT 616
 OY 635 TCTCCAGAGCCCAAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 694
 DB 617 GCTCCGGGAGCTTACAAAG 676
 OY 695 AGTGGCTGACACCCGCTGCGCAATGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 754
 DB 677 ACATGAGGAGCAATGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 736
 OY 755 GAGAGTACGATGGGGGTGAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 814
 DB 737 GCGAGTTTCATGAG 796
 OY 815 CCACCTTGGGCTGAGAGCTTCACAGCCCAACACTGTGAGAACTTGTCCAGAGCTGG 874
 DB 797 CAACACAGAGACTGAGTGCATGATGCTCAATCAACAGAGAGAGAGAGAGAGAGAGAG 856
 OY 875 CCAAGGCAACCGGCTGCTGCTATCTCCCTCCACAGAGCTGCTGACATCTTCAGGC 934
 DB 857 CTCGAGGAGAGCAATGATGATGCTCAACATCCACAGAGAGAGAGAGAGAGAGAGAG 916
 OY 935 TGTTCATCTGCTCTCCGATGAGCTGAGACCCCATCTACTTATGAGGCGGCCAGC 994
 DB 917 ACTTCGACAAATTCCTCCATCTGATTCAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 976
 OY 995 ACATGCTCAGATATTCACAGCCATCGCTACCCCTGCTGCTACAGCAATCTGCTG 1054
 DB 977 AGATGCTGCTGCTTCAATACGTGTGTTACCCCTTCTGCAATTCATCTGCTTGG 1036
 OY 1055 ACTTCATGATGAGCTGACAGCAATTCAGAGGCGCAGAGAGAGAGAGAGAGAGAGAG 1114
 DB 1037 ATTTTACATGAGCTGATGATCAGTGAGACCCCAAGAGAGAGAGAGAGAGAGAGAG 1096
 OY 1115 GGGAGAGGCTCAGTCACTGCCA 1137
 DB 1097 ACAAGCGAGTACAGATGCTGGA 1119

RESULT 15
 ABR51687
 ID ABR51687 standard; cDNA; 1069 BP.
 AC ABR51687;
 XX
 DT 30-JUL-2002 (first entry)
 XX
 DE cDNA encoding hamster ABCG5 protein.
 XX
 KW Hamster: ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW ss.
 XX
 OS Cricetinae sp.
 XX
 FH Key
 FT Location/Qualifiers
 FT 30..1049
 FT /*tag- a
 FT /partial-
 FT /product- "Hamster ABCG5 protein"
 FT /note- "This sequence lacks both a start and stop codon"
 XX
 PN W0200227016-A2.
 XX
 PD 04-APR-2002.
 XX
 PE 25-SEP-2001; 2001WO-US29859.
 XX
 PR 25-SEP-2000; 2000US-235268P.
 XX
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 XX
 PI Patel SB, Dean M;
 DR WPI; 2002-416483/44.
 DR P-PDB; AAU96987.
 XX
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 XX
 PS Example 3; Page 47; 66pp; English.
 XX
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or in mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present nucleic
 CC acid sequence encodes the hamster ABCG5 protein of the invention.
 XX
 SO Sequence 1069 BP; 266 A; 282 C; 273 G; 248 T; 0 other;

Query Match 6.5%; Score 173.2; DB 24; Length 1069;
 Best Local Similarity 54.1%; Pred. No. 4,1e-36;

	Matches	397;	Conservative	0;	Mismatches	333;	Indels	4;	Gaps	2;
OY	418	TCAGCTTGAGGAGAGCCTCTCTAGATGTGATCAGTGGCCGAGGTACAGGGGCAAG	477							
Db	1	TCAGGCTCAGGAGAAACACAGTTGCT-6GTGCACTCTCCGGAGGCTCGACGACAGGG	59							
OY	478	ATCAAGTAGGCCAGATCTGGATCAATGAGGACGCCAGCTGCTCAGCTGGTGAAG	537							
Db	60	ACCCGGAAGGGGAGGTCTTGTGAAGCCCGTGAAGTGGCAGAGGACCATGTCACAGC	119							
OY	538	TGTGTGGCCACGTGGCCGACACACAGCAGTCTCCCAACTGACTGTGCGAGAGACC	597							
Db	120	TGCTTCTCTATGTCTGACAGAGGACGCTCTTCTGACAGCTCTCAGCTGAGAGAGC	179							
OY	598	TTGGCTTCATTGCTCCAGATCTGGCTGGCCAGAACCTCTCCAGCCGACAGTACAA	657							
Db	180	CTGGGCTACAGGGGAGTGTGCTGCTCCGCTCCGAGTACTCTTC--GGACTTATGACAA	236							
OY	658	AGGTTGAGAGACGTGATCCGAGCTGGAGCTTGAAGCAGTGTGACACCCGGGTGGC	717							
Db	237	AAGTAGAGGACATGAGAAAGACTAAGTGTGACCCAGCGCAGACCCAGATATTTGC	296							
OY	718	AACATGTACGTGGGGGCTGTGCGGGGTGAGCGCAGAGAGTACCATTTGGGGTCAAG	777							
Db	297	AACTATATTTTGGGGAAATTTCCAGTGGCGAGCGCGCGAGTCTCATGCGAGCCAA	356							
OY	778	CTCCTGTGAACCCAGGAATCCTATCTGAGACAAACCCACTTGGGCTGACAGCTTC	837							
Db	357	CTCATATGAGAACCCAGAAATCATGATTTTGTATGAGCAACACAGACGACTGACATG	416							
OY	838	ACAGCCCAACACTGTGTGAAGACCTTGTCCAGCTGCAAAAGGACCGGCTGGTCTC	897							
Db	417	ACTGCAAAATCAAAATTCATCTCCCTGAGAGAGCTGCTGCGAGGACCCCATTTGATC	476							
OY	898	ATCTCCCTCCACCAAGCTGCTGTGACATCTTCAAGCTGTTTATGTGTCTCTGATG	957							
Db	477	GTCACCATCCACAGCGCTGCTGAGCTCTTCAACACTTGCACAAATTTGCAATCTG	536							
OY	958	ACGCTGGACACCCCAATCTAGAGGGGGGCCACACATGTGTCCAGTATTTCACAGCC	1017							
Db	537	ACTTAGGAGAGATGTGTTCTTGTGCAACCCGAGAGAAATGCTGCACTTCAATPAGC	596							
OY	1018	ATCGGCTACCCCTGCTCTGCTACAGCAATCTGCTGACTTATGTGAGACCTGACAGC	1077							
Db	597	TGTGTATACCTTGTCTCAACATTCACACCCCTTGTGACTTCACTGACCTTGACATCA	656							
OY	1078	ATTGACAGGCGAGCAGAGAGAGCAAGTATGCGCCACAGGAGAGGCTCAGTCACTGCA	1137							
Db	657	GTTGATACCCAGAGCAGAGAGGAGAAATAGAAACCTACAGAGAGTCCAGATGCTGAA	716							
OY	1138	GCCCTGTTCTTAGA 1151								
Db	717	TCTGCTTCAGAGA 730								

Search completed: July 27, 2003, 00:55:37
 Job time : 650.803 secs

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

(without alignments)
11512.153 Million cell updates/sec

sequence: 1 gtgtccctgtctccaggaac.....caatlaaatgtattgac 2669

Searched: 2888711 seqs, 20454813386 residues

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

GenEmb1:*

3:	gb_in:*
2:	gb_mngt:*
1:	em_hlgo:*
4:	gb_om:*
5:	gb_ov:*
6:	gb_pat:*
7:	gb_ph:*
8:	gb_pl:*
9:	gb_pr:*
10:	gb_ro:*
11:	gb_sts:*
12:	gb_sy:*
13:	gb_un:*
14:	gb_vl:*
15:	em_ba:*
16:	em_fun:*
17:	em_hm:*
18:	em_in:*
19:	em_jn:*
20:	em_om:*
21:	em_ot:*
22:	em_ov:*
23:	em_pat:*
24:	em_pln:*
25:	em_pl:*
26:	em_ro:*
27:	em_sts:*
28:	em_un:*
29:	em_vl:*
30:	em_hlg_hum:*
31:	em_hlg_inv:*
32:	em_hlg_other:*
33:	em_hlg_mus:*
34:	em_hlg_rod:*
35:	em_hlg_rod:*
36:	em_hlg_mam:*
37:	em_hlg_vrt:*
38:	em_sy:*
39:	em_hlgo_hum:*
40:	em_hlgo_mus:*
41:	em_hlgo_other:*

Pred. No. is the number of results predicted by chance to have a

Result	Score	Match	Query	Length	DB	ID	Description
1	2669	100.0	2669	6	AX685735	AX685735 Sequence	
2	2655.2	99.5	2679	6	AF324494	AF324494 Homo sapi	
3	2020.4	75.7	2022	9	AF320294	AF320294 Homo sapi	
4	1680.6	63.0	3339	6	AX478099	AX478099 Sequence	
5	1518.6	56.9	4829	10	AF351785	AF351785 Rattus no	
6	1511.6	55.6	3674	10	AF324495	AF324495 Mus muscu	
7	1454.2	54.5	2284	10	AY196216	AY196216 Mus muscu	
8	1449.4	54.3	2285	10	AY196215	AY196215 Mus muscu	
9	1430	53.6	2019	6	AX685731	AX685731 Sequence	
10	724	27.1	127066	9	AC084265	AC084265 Homo sapi	
11	724	27.1	182261	2	AC087053	AC087053 Homo sapi	
12	722.4	27.1	139342	9	AC108476	AC108476 Homo sapi	
13	663.6	24.9	2201	9	AF31812933	AF31812933 Homo sapi	
14	274.2	10.3	660	9	AF31812806	AF31812806 Homo sapi	
15	273.4	10.2	1292	9	AF35182251	AF351822 Homo sapi	
16	260.6	9.8	68166	2	AC084712	AC084712 Homo sapi	
17	244	9.1	1323	9	AF351812504	AF351815 Homo sapi	
18	231.2	8.7	250	11	G29195	G29195 human STS S	
19	230.4	8.6	237445	2	AC120701	AC120701 Rattus no	
20	230.4	8.6	312858	2	AC112747	AC112747 Rattus no	
21	229.4	8.6	40929	10	AY145899	AY145899 Rattus no	
22	226.8	8.5	1387	10	AF35179506	AF351804 Mus muscu	
23	207.4	7.8	156829	2	AC120532	AC120532 Oryza sat	
24	204.8	7.7	1378	2	AF351799511	AF351809 Mus muscu	
25	204.8	7.7	185545	2	AC122243	AC122243 Mus muscu	
26	203.8	7.6	685	9	AF351812509	AF351820 Homo sapi	
27	203.6	7.6	1920	6	AX456519	AX456519 Sequence	
28	203.6	7.6	2340	6	AX320883	AX320883 Sequence	
29	203.6	7.6	2340	6	AX685733	AX685733 Sequence	
30	203.6	7.6	2340	6	AF320293	AF320293 Homo sapi	
31	203.6	7.6	2516	6	AX456520	AX456520 Sequence	
32	203.6	7.6	2740	9	AF312715	AF312715 Homo sapi	
33	202.4	7.6	400	11	G16507	G16507 human STS S	
34	195	7.3	2351	10	AX196873	AX196873 Mus muscu	
35	194.4	7.3	2035	6	AX456526	AX456526 Sequence	
36	194.4	7.3	2470	10	AF312714	AF312714 Rattus no	
37	194	7.3	2351	10	AY195872	AY195872 Mus muscu	
38	193.4	7.2	1915	6	AX456523	AX456523 Sequence	
39	193.4	7.2	1959	6	AX685729	AX685729 Sequence	
40	193.4	7.2	2258	6	AX320881	AX320881 Sequence	
41	193.4	7.2	2354	6	AX456524	AX456524 Sequence	
42	193.4	7.2	2354	10	AF312713	AF312713 Mus muscu	
43	190.4	7.1	1716	10	AF351799514	AF351811 Mus muscu	
44	184.2	6.9	1470	10	AF351799503	AF351802 Mus muscu	
45	173.2	6.5	1069	6	AX456528	AX456528 Sequence	

ALIGNMENTS

RESULT 1
LOCUS AX685735
DEFINITION Sequence 7 from Patent WO02081691.
ACCESSION AX685735
VERSION AX685735.1 GI:29371744
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS Hobbs,H.H., Shan,B., Barnes,R. and Tian,H.
TITLE Adc95 and adc98: compositions and methods of use
JOURNAL Patent: WO 02081691-A 7 17-OCT-2002.

PAT 29-MAR-2003

660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688
 689
 690
 691
 692
 693
 694
 695
 696
 697
 698
 699
 700
 701
 702
 703
 704
 705
 706
 707
 708
 709
 710
 711
 712
 713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728
 729
 730
 731
 732
 733
 734
 735
 736
 737
 738
 739
 740
 741
 742
 743
 744
 745
 746
 747
 748
 749
 750
 751
 752
 753
 754
 755
 756
 757
 758
 759
 760
 761
 762
 763
 764
 765
 766
 767
 768
 769
 770
 771
 772
 773
 774
 775
 776
 777
 778
 779
 780
 781
 782
 783
 784
 785
 786
 787
 788
 789
 790
 791
 792
 793
 794
 795
 796
 797
 798
 799
 800
 801
 802
 803
 804
 805
 806
 807
 808
 809
 810
 811
 812
 813
 814
 815
 816
 817
 818
 819
 820
 821
 822
 823
 824
 825
 826
 827
 828
 829
 830
 831
 832
 833
 834
 835
 836
 837
 838
 839
 840
 841
 842
 843
 844
 845
 846
 847
 848
 849
 850
 851
 852
 853
 854
 855
 856
 857
 858
 859
 860
 861
 862
 863
 864
 865
 866
 867
 868
 869
 870
 871
 872
 873
 874
 875
 876
 877
 878
 879
 880
 881
 882
 883
 884
 885
 886
 887
 888
 889
 890
 891
 892
 893
 894
 895
 896
 897
 898
 899
 900
 901
 902
 903
 904
 905
 906
 907
 908
 909
 910
 911
 912
 913
 914
 915
 916
 917
 918
 919
 920
 921
 922
 923
 924
 925
 926
 927
 928
 929
 930
 931
 932
 933
 934
 935
 936
 937
 938
 939
 940
 941
 942
 943
 944
 945
 946
 947
 948
 949
 950
 951
 952
 953
 954
 955
 956
 957
 958
 959
 960
 961
 962
 963
 964
 965
 966
 967
 968
 969
 970
 971
 972
 973
 974
 975
 976
 977
 978
 979
 980
 981
 982
 983
 984
 985
 986
 987
 988
 989
 990
 991
 992
 993
 994
 995
 996
 997
 998
 999
 1000

D	601	GCCTTATGTCGCCAGATGGGGGCTGGCCAGAACCTTCTCCAGGCCCAAGCTGACAAAGG	660
Q	661	GTGGAGAGCTGATGCGCGGAGCTGCGGCTTAGGCACTGCGGTACACCCGCGTGGCAAC	720
D	661	GTGGAGAGCTGATGCGCGGAGCTGCGGCTTAGGCACTGCGGTACACCCGCGTGGCAAC	720
Q	721	ATGTAGTGGCGGGGGGTGTGCGGGGGGTGAGCGCAGAGAGTCAAGCTTGGGGTGCAGCTC	780
D	721	ATGTAGTGGCGGGGGGTGTGCGGGGGGTGAGCGCAGAGAGTCAAGCTTGGGGGTGCAGCTC	780
Q	781	CTGTGGAAACCCAGGAATCCCTTAATTCGAGAAACCCACCTCGGGGCTGCACAGCTTCACA	840
D	781	CTGTGGAAACCCAGGAATCCCTTAATTCGAGAAACCCACCTCGGGGCTGCACAGCTTCACA	840
Q	841	GCCCAACAACCTGGTGAAGAAGCTTGTCCAGGCTGGCCAAAGGCAACCGGCTGGTGCATAC	900
D	841	GCCCAACAACCTGGTGAAGAAGCTTGTCCAGGCTGGCCAAAGGCAACCGGCTGGTGCATAC	900
Q	901	TCCCTTCACACAGCCTCGCTGCTGACATCTTTCAGGCTGTTTGATCTGGTCTCTGTATGAG	960
D	901	TCCCTTCACACAGCCTCGCTGCTGACATCTTTCAGGCTGTTTGATCTGGTCTCTGTATGAG	960
Q	961	TCGTGGACCCCACTACTATTAGGGGGGGGCCAGACATAGTGGTCCAGATTTTCACAGCATAC	1020
D	961	TCGTGGACCCCACTACTATTAGGGGGGGGCCAGACATAGTGGTCCAGATTTTCACAGCATAC	1020
Q	1021	GGCTACCCCTGTCTCTGCTACAGCAATCTGCTGACTTCTATGTGGACCTGCACAGCAT	1080
D	1021	GGCTACCCCTGTCTCTGCTACAGCAATCTGCTGACTTCTATGTGGACCTGCACAGCAT	1080
Q	1081	GACAGGCGCAGCAGAGAGCAGAGAAATTGGCCACAGGAGAGGCTCAGTCACTGCGAGCC	1140
D	1081	GACAGGCGCAGCAGAGAGCAGAGAAATTGGCCACAGGAGAGGCTCAGTCACTGCGAGCC	1140
Q	1141	CGGTTCTGTGAAGAAAGTGGCTGACTTAATGACTTTCTATGAAAGCAGAGAGAGAT	1200
D	1141	CGGTTCTGTGAAGAAAGTGGCTGACTTAATGACTTTCTATGAAAGCAGAGAGAGAT	1200
Q	1201	CTTGAAGAGACAGCTGTGTGAAAGACAGCTGAGCCCACTGAGACACCACTGCGCCCG	1260
D	1201	CTTGAAGAGACAGCTGTGTGAAAGACAGCTGAGCCCACTGAGACACCACTGCGCCCG	1260
Q	1261	AGTCTTACGAGAATGCTGTGGGGGGGTGCAGCAGTTTACAGCCTGATCCGTGCAGATT	1320
D	1261	AGTCTTACGAGAATGCTGTGGGGGGGTGCAGCAGTTTACAGCCTGATCCGTGCAGATT	1320
Q	1321	TCCAAAGACTTCCGAGACCTGCGCAACCCCTCATTCATCANTGGGGCGGAGGCCCTGTGATG	1380
D	1321	TCCAAAGACTTCCGAGACCTGCGCAACCCCTCATTCATCANTGGGGCGGAGGCCCTGTGATG	1380
Q	1381	TCAATGAGCAGTGGGCTGCTCTATTATTGGCCATGAGAGCATCAGCTCCTTCATGAGAT	1440
D	1381	TCAATGAGCAGTGGGCTGCTCTATTATTGGCCATGAGAGCATCAGCTCCTTCATGAGAT	1440
Q	1441	ACAGCGGCCCTCTTGTCTCATGATGCGTGTCTCATCCCTTCACAGCTATCTGTGATGATC	1500
D	1441	ACAGCGGCCCTCTTGTCTCATGATGCGTGTCTCATCCCTTCACAGCTATCTGTGATGATC	1500
Q	1501	ATCTCAAAATGTTACTCAGAGAGGGCAATGCTTACTATGAACTGGAAGAGGGCTGTAC	1560
D	1501	ATCTCAAAATGTTACTCAGAGAGGGCAATGCTTACTATGAACTGGAAGAGGGCTGTAC	1560
Q	1561	ACCACTGGGCCATTTCTTGGCCAAAGATCCGCGGGAGGCTTCGGGACACATGTCGCCATC	1620
D	1561	ACCACTGGGCCATTTCTTGGCCAAAGATCCGCGGGAGGCTTCGGGACACATGTCGCCATC	1620
Q	1621	ATCATCATCTACGGGATGCCACCACTACTGCTGCGCAACCTGAGGCCAGGCTCCAGGCC	1680
D	1621	ATCATCATCTACGGGATGCCACCACTACTGCTGCGCAACCTGAGGCCAGGCTCCAGGCC	1680
Q	1681	TTCCGCTGCACTTCTCTGCTGTGGGTGGGTGGTGGTCTTCTGTGACGAGATTATGGCCCTG	1740
D	1681	TTCCGCTGCACTTCTCTGCTGTGGGTGGGTGGTGGTCTTCTGTGACGAGATTATGGCCCTG	1740

```

1741 GCCGCCGCCGCTCTCCACCTTCACATGCGCTCTCTCTGAGCAATGCCCTCTAC
1741 GCGCGCGCGGCGCTGCTCCGACCTTCACATGCGCTCTCTCTGAGCAATGCCCTCTAC
1801 AACTCTCTCTACCTGCGCGGCGCTTCATGATTAACCTTGAGACAGCCCTGACAGTCC
1801 AACTCTCTCTACCTGCGCGGCGCTTCATGATTAACCTTGAGACAGCCCTGACAGTCC
1861 GCGTGAATTTCCAAAGTCTCTCTGCGGCGTGTGTTGAGGCGCTGATGAAGTTCAG
1861 GCGTGAATTTCCAAAGTCTCTCTGCGGCGTGTGTTGAGGCGCTGATGAAGTTCAG
1921 TTCACAGAGAACTTAAATGAGCTCTGCGGAGACCTCACCATGCGGCTCTCAGAGAT
1921 TTCACAGAGAACTTAAATGAGCTCTGCGGAGACCTCACCATGCGGCTCTCAGAGAT
1981 AAAATCTCAGTGCATGAGCTGAGCTGATACCTCTCTACGCGCATCTACCTCATCTC
1981 AAAATCTCAGTGCATGAGCTGAGCTGATACCTCTCTACGCGCATCTACCTCATCTC
2041 ATTGCGCTCAGCGGCTCTCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
2041 ATTGCGCTCAGCGGCTCTCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
2101 AAACCAAGTCAAGAGCTGATGATTCACGCGCAGACGCTCTGCGGCGGAGAGCTGAGC
2101 AAACCAAGTCAAGAGCTGATGATTCACGCGCAGACGCTCTGCGGCGGAGAGCTGAGC
2161 AGACCTTCAACTGACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
2161 AGACCTTCAACTGACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
2221 CCTTACAGATGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
2221 CCTTACAGATGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
2281 GATGCGAGTAAATTAAGACAGTGCAGAAAGGATTTCTGCTCTCTCTCTCTCTCTCTCTCT
2281 GATGCGAGTAAATTAAGACAGTGCAGAAAGGATTTCTGCTCTCTCTCTCTCTCTCTCTCT
2341 GATGCGAGTAAATTAAGACAGTGCAGAAAGGATTTCTGCTCTCTCTCTCTCTCTCTCTCT
2341 GATGCGAGTAAATTAAGACAGTGCAGAAAGGATTTCTGCTCTCTCTCTCTCTCTCTCTCT
2401 TATGCAATTTATATGAGCACTGATATGATGAGGAGCAAACTGAGAAATGGAATGGTAG
2401 TATGCAATTTATATGAGCACTGATATGATGAGGAGCAAACTGAGAAATGGAATGGTAG
2461 TATGCAATTTATATGAGCACTGATATGATGAGGAGCAAACTGAGAAATGGAATGGTAG
2461 TATGCAATTTATATGAGCACTGATATGATGAGGAGCAAACTGAGAAATGGAATGGTAG
2521 GCTTCAATCTTTCAGAGGCGCCACACTCCGCTGCTGAGGAGCAATCAATCAGAAAGTACC
2521 GCTTCAATCTTTCAGAGGCGCCACACTCCGCTGCTGAGGAGCAATCAATCAGAAAGTACC
2581 GCTTCAATCTTTCAGAGGCGCCACACTCCGCTGCTGAGGAGCAATCAATCAGAAAGTACC
2581 GCTTCAATCTTTCAGAGGCGCCACACTCCGCTGCTGAGGAGCAATCAATCAGAAAGTACC
2641 CAACGTGAACAATTAATTAATGATTTAGC 2669
2641 CAACGTGAACAATTAATTAATGATTTAGC 2669

```

```

SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Lu, K., Lee, M.-H., Hazard, S., Brookes-Wilson, A., Hida, K., Kojima, H.,
Osse, L., Stalenhoef, A.F., Mettenen, T., Bjorkhem, I., Bruckert, E.,
Pandya, A., Brewer, H.B. Jr., Salen, G., Dean, M., Sirtava, A. and
Patel, S.B.
TITLE Two genes that map to the STSL locus cause sitosterolemia: genomic
structure and spectrum of mutations involving sterol-1 and
sterol-2, encoded by ABCG5 and ABCG8, respectively
JOURNAL Am. J. Hum. Genet. 69 (2), 278-290 (2001)
MEDLINE 21344600
PUBMED 11452359
REFERENCE 2 (bases 1 to 2679)
AUTHORS Lu, K., Lee, M.-H. and Patel, S.B.
JOURNAL Direct Submission
TITLE Submitted (29-NOV-2000) Division of Endocrinology, Diabetes and
Medical Genetics, Medical University of South Carolina, 114 Doughty
Street, STB541, Charleston, SC 29403, USA
FEATURES
source
1..2679
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2p21: between D2S2294 and D2S2298"
/tissue_type="liver"
1..2679
/gene="ABCG8"
91..2112
/gene="ABCG8"
/codon_start=1
/product="sterolin-2"
/protein_id="AAK84078.1"
/db_xref="GI:15088540"
/translation="MAGRAAEERGLPKGATPDNSGLDRLFSESDNSLYFTYSGOP
NLEVRDNCVDLADVPMELOAFEMPTSPSCSCGLIONLSFKYRSGMLA
IIGSGGCRALDLVITGRGGRKIKSGQIVINQSPOLVRKCAVAVRQHNOLPN
LTVRETLAFIOMRLPRTFEOAORDEVIARLRQCADTVRGNNYVRLSGGER
RVSIVDVLINPCTIILDEPTSLDSTAHNLKTLRLKRNLYLSLQPRSDI
PRLDIVILMNSGPPIYLGAOHVOTFTAIQPCPRTSNADTVLDSIDRSRQ
ELATREKQSLAALFLEKVRDLDFLHGAERKLDDEDCVSVPLDNLPSPTKM
PGAVQOFTLLIRROISNDFRDLPLHLHGAERKLDDEDCVSVPLDNLPSPTKM
LLFMIGALIPFNVLIDVSKYSEKRAMLYELEGLVTTGYFPFAKILGEPEHCAVI
IIVGPTWYMLANLRLPGLDFLHLFLWLVVFCGRLMAAALLPHEMPSNAL
YNSPYLAGGEMINISLMTVPAMISKYSFLMPCREGELMKLIDPSRTYKMPKGNLTIAV
SGDKILSMEIDSLPLAIVLIVIGLSGFEVLYVSLRFTKRPQSDW"
BASE COUNT 609 a 767 c 720 g 583 t
ORIGIN
Query Match 99.5%; Score 2655.2; DB 9; Length 2679;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2657; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
10 CTCGAGAAACAGAGTGAAGACACTGGCCCTGCGCAGCAGCAGCTGGCTTAAGAGACT 69
1 CTCGAGAAACAGAGTGAAGACACTGGCCCTGCGCAGCAGCAGCTGGCTTAAGAGACT 60
70 GCACCCGAGGCTCAGACCTGTGTGGCCCATGCGCGGGAAGCGCGCAGAGAGAGAGG 129
61 GCACCCGAGGCTCAGACCTGTGTGGCCCATGCGCGGGAAGCGCGCAGAGAGAGAGG 120
130 CTGCGGAAAGGCGCAGTCCCGCAGATACCTCGGCGCTCCAGGATATCTTCTCTCT 189
121 CTGCGGAAAGGCGCAGTCCCGCAGATACCTCGGCGCTCCAGGATATCTTCTCTCT 180
190 GAAAGTGAACAACACCTGTACTTACCTACAGTGGCGCAGCCACACCTGGAGGTGAGA 249
181 GAAAGTGAACAACACCTGTACTTACCTACAGTGGCGCAGCCACACCTGGAGGTGAGA 240
250 GACCTCAACTACACAGTGGAGCTGCGCTCTCAGAGTCCCTTGTGAGACAGCTGCTCAG 309

```

```

RESULT 2
LOCUS AF324494
DEFINITION Homo sapiens sterol-2 (ABCG8) mRNA, complete cds.
ACCESSION AF324494
VERSION AF324494.1 GI:15088539
KEYWORDS

```

Db	241	GNACCTCACTGCCAGGTGGACCTGGCCCTTCAGTCCCTTGGTTTAGCAAGCTGGCTCAG	300
OY	310	TTCAAGATGCCCTGGACATCTCCACAGTCCAGAAATCTTGTGAGCTGGGCATCCAGAAC	369
Db	301	TTCAAGATGCCCTGGACATCTCCACAGTCCAGAAATCTTGTGAGCTGGGCATCCAGAAC	360
OY	370	CTAAGCTTAAAGTGAAGTGGGCGAGATGCTGTCGCTATCAATYAGGAGCTTCAGTGTGTGG	429
Db	361	CTAAGCTTAAAGTGAAGTGGGCGAGATGCTGTCGCTATCAATYAGGAGCTTCAGTGTGTGG	420
OY	430	AGAGCCTCTGTGCTATGATGATCAGTGGCCGAGGTCAGAGGGGCAAGATCAAGTCAAGGC	489
Db	421	AGAGCCTCTGTGCTATGATGATCAGTGGCCGAGGTCAGAGGGGCAAGATCAAGTCAAGGC	480
OY	490	CAGATCTGATCAATATGGGAGGCCACAGCTTGCCTCAGCTGATGAGAAAGTGTGGGCCAC	549
Db	481	CAGATCTGATCAATATGGGAGGCCACAGCTTGCCTCAGCTGATGAGAAAGTGTGGGCCAC	540
OY	550	GTGCGCCACACACACACAGCTGCTCCCAACTTGTACTGTGCGAGAGACCTTGGCTCTCAT	609
Db	541	GTGCGCCACACACACACAGCTGCTCCCAACTTGTACTGTGCGAGAGACCTTGGCTCTCAT	600
OY	610	GGCCAGATGCGGCTGGCCAGAACCTTCTCCCAAGGCCACAGCTGACAAAAGGTTGAGAGAC	669
Db	601	GGCCAGATGCGGCTGGCCAGAACCTTCTCCCAAGGCCACAGCTGACAAAAGGTTGAGAGAC	660
OY	670	GTGATCGCGAGACTCGGCGCTTAGGAGAGTGCCTGACACCCGGGTGGGCAACATGTACGAG	729
Db	661	GTGATCGCGAGACTCGGCGCTTAGGAGAGTGCCTGACACCCGGGTGGGCAACATGTACGAG	720
OY	730	CGGGGGTGTGCGGGGGGTGAGCGCAGAGAGTCAAGCATYGGGGTGCAGCTCTCTGTGGAAC	789
Db	721	CGGGGGTGTGCGGGGGGTGAGCGCAGAGAGTCAAGCATYGGGGTGCAGCTCTCTGTGGAAC	780
OY	790	CGAGAAATCCTAATCTGAGCAACCCACCTTGGGCTGTGACAGCTTACAGCCCAACAC	849
Db	781	CGAGAAATCCTAATCTGAGCAACCCACCTTGGGCTGTGACAGCTTACAGCCCAACAC	840
OY	850	CTGTGTGAAGACCTTGTCCAGGCTGTGGCCAAAAGGCAACCGGCTGGTCAATCTCCCTCCAC	909
Db	841	CTGTGTGAAGACCTTGTCCAGGCTGTGGCCAAAAGGCAACCGGCTGGTCAATCTCCCTCCAC	900
OY	910	CAGCCTCGCTTGACATCTTCAAGGCTGTTTGAATCTGTGCTTCCTGATGACGTCTGGCAC	969
Db	901	CAGCCTCGCTTGACATCTTCAAGGCTGTTTGAATCTGTGCTTCCTGATGACGTCTGGCAC	960
OY	970	CCCATCTACTTAGGGGGCGGCCACAGCAATGGTCCAGTATTTTCACGCCATGGGCTACCCC	1029
Db	961	CCCATCTACTTAGGGGGCGGCCACAGCAATGGTCCAGTATTTTCACGCCATGGGCTACCCC	1020
OY	1030	TGTCCTCCCTACAGCAATCTGTGCTACTTCTATGTGGACCTGACCAAGATTTGACAGGGGC	1089
Db	1021	TGTCCTCCCTACAGCAATCTGTGCTACTTCTATGTGGACCTGACCAAGATTTGACAGGGGC	1080
OY	1090	AGCAGAGACAGGAATTTGGCCACAGGAGAGAAGGCTCAGTCACTTCGACGCCCTCTTCTTA	1149
Db	1081	AGCAGAGACAGGAATTTGGCCACAGGAGAGAAGGCTCAGTCACTTCGACGCCCTCTTCTTA	1140
OY	1150	GAAAAAGTGGCTGATTTAGATGACTTTTATGTGAAAGCAGAGACGAAAGATCTTGACGAG	1209
Db	1141	GAAAAAGTGGCTGATTTAGATGACTTTTATGTGAAAGCAGAGACGAAAGATCTTGACGAG	1200
OY	1210	GACACCTGTGTGAAAGCAGGCTGACCCACATAGACACCAACTGCTCCGACAGTCTTCACG	1269
Db	1201	GACACCTGTGTGAAAGCAGGCTGACCCACATAGACACCAACTGCTCCGACAGTCTTCACG	1260
OY	1270	AAGATGCTGTGGGGCGGTGACAGATTTTACAGAGCTGATCCGTCTGATTTTCAACAGAC	1329
Db	1261	AAGATGCTGTGGGGCGGTGACAGATTTTACAGAGCTGATCCGTCTGATTTTCAACAGAC	1320
OY	1330	TTCCGAGACCTTGGCCACCTCTCTCATCTATGGGGGAGAGGCTGTCTGATGTCAATGACC	1389

Db	1321	TTCCGAGACCTGCCCCACCCCTCCYCATCCATG6GGGGAAGGCCCTGTCTATGTCAATGAC	1387
OY	1390	ATCGGCTTCCTCATTTTGGCCATGAGGAGCATCGACGTCTCTCATGATGATACAGCCGC	1445
Db	1381	ATCGGCTTCCTCATTTTGGCCATGAGGAGCATCGACGTCTCTCATGATGATACAGCCGC	1440
OY	1450	CTCTTGTTTCATGATGCGGTCTCATACCCCTTTCACAGCTCATCTTGATGATCATCTCCAAA	1509
Db	1441	CTCTTGTTTCATGATGCGGTCTCATACCCCTTTCACAGCTCATCTTGATGATCATCTCCAAA	1500
OY	1510	TCTTACTCAGAGAGGGCAATGCTTTTACTATGAACTGGAAGACGGGCTGTACACCACTGGT	1565
Db	1501	TCTTACTCAGAGAGGGCAATGCTTTTACTATGAACTGGAAGACGGGCTGTACACCACTGGT	1560
OY	1570	CCATATTTCTTTGGCAAGATCTCTGGGGAGCTTCGGGAGCACTGTGCTTACATCATCATC	1620
Db	1561	CCATATTTCTTTGGCAAGATCTCTGGGGAGCTTCGGGAGCACTGTGCTTACATCATCATC	1620
OY	1630	TACGGGATGCCCCACCTACGTGGGTGGCCAAACCTGAGGCCAGGCTTCACGCCCTTCTGTCG	1688
Db	1621	TACGGGATGCCCCACCTACGTGGGTGGCCAAACCTGAGGCCAGGCTTCACGCCCTTCTGTCG	1688
OY	1690	CACCTTCCTGCTGCTGTGGCTGGGTGCTTCTGTGTGACAGAAATATGCCCCGCGCGCG	1749
Db	1681	CACCTTCCTGCTGCTGTGGCTGGGTGCTTCTGTGTGACAGAAATATGCCCCGCGCGCG	1748
OY	1750	GCCCGCTCCGCCACCTTCACATGGCCCTCTTCTCAGACAAATGCCCTTACAACTCCTTC	1809
Db	1741	GCCCGCTCCGCCACCTTCACATGGCCCTCTTCTCAGACAAATGCCCTTACAACTCCTTC	1800
OY	1810	TACCTCGCGCGGGGCTTCATGATGAACCTTGAAGACAGCCGTGGACAGTCCCGCGGGATTT	1865
Db	1801	TACCTCGCGCGGGGCTTCATGATGAACCTTGAAGACAGCCGTGGACAGTCCCGCGGGATTT	1860
OY	1870	TCCAAAGTCTCCTTCTCTGCGGTGTGTTTGAAGGGCTGATGAAGATTCACCTCAGCAGA	1929
Db	1861	TCCAAAGTCTCCTTCTCTGCGGTGTGTTTGAAGGGCTGATGAAGATTCACCTCAGCAGA	1920
OY	1930	AGAACTTATTAATAAAGCCCTCTCGGGAACTCAGCATCGCGGTCTCAGAGATTAATAATCTC	1989
Db	1921	AGAACTTATTAATAAAGCCCTCTCGGGAACTCAGCATCGCGGTCTCAGAGATTAATAATCTC	1980
OY	1990	AGTGCATGAGAGCTGAGTCTGATACCCCTTCATACGCATCTACCTCATGATTTGGGCTC	2049
Db	1981	AGTGCATGAGAGCTGAGTCTGATACCCCTTCATACGCATCTACCTCATGATTTGGGCTC	2040
OY	2050	AGCGGTGCTTCATGCTGTCTGTACTAGCTGTCTTAAGTTCATCAAAACAGAAACCAAGT	2109
Db	2041	AGCGGTGCTTCATGCTGTCTGTACTAGCTGTCTTAAGTTCATCAAAACAGAAACCAAGT	2100
OY	2110	CAGACGTGTATTCAGGCCAGACGTGTGCCGCTGTGGGGACCTGAGCAGACCCCTTC	2169
Db	2101	CAGACGTGTATTCAGGCCAGACGTGTGCCGCTGTGGGGACCTGAGCAGACCCCTTC	2160
OY	2170	AACGTGCACTCCCTCCTCAGAGAGCCCTTCCTCGGGGAGACAGTAGAGACAAATGACCTTACAGA	2229
Db	2161	AACGTGCACTCCCTCCTCAGAGAGCCCTTCCTCGGGGAGACAGTAGAGACAAATGACCTTACAGA	2220
OY	2230	TGCTCAGCTACATCCGGGCCAGGGTGTCTGACGTGGCAGACAGCCAGCAGAGATGGCAGT	2289
Db	2221	TGCTCAGCTACATCCGGGCCAGGGTGTCTGACGTGGCAGACAGCCAGCAGAGATGGCAGT	2280
OY	2290	AGAAATTAAGACAGTTCGAAAGGAGATTTTGTCTCACTGGCAGAGAGACTGCGATGACTGGAG	2349
Db	2281	AGAAATTAAGACAGTTCGAAAGGAGATTTTGTCTCACTGGCAGAGAGACTGCGATGACTGGAG	2340
OY	2350	AAAACCTGCACTCGGTGGCAGCTTCACAGCGTTGCAATTTTTCCTTTTATATGCAATTT	2409
Db	2341	AAAACCTGCACTCGGTGGCAGCTTCACAGCGTTGCAATTTTTCCTTTTATATGCAATTT	2400
OY	2410	ATATAGCAACTCGATATAGAGATGAGAGCAAACTAGGAATGAATTTGGGTAGCTAGACTGT	2469
Db	2401	ATATAGCAACTCGATATAGAGATGAGAGCAAACTAGGAATGAATTTGGGTAGCTAGACTGT	2460

OY 2470 GCAGGATTTGTTGAACTGGAGGAGCAATTAACAGTAGACAGATTGGTTCATCT 2529
Db 2461 GCAGGATTTGTTGAACTGGAGGAGCAATTAACAGTAGACAGATTGGTTCATCT 2520
OY 2530 TCAGAGGGCCCCCACTCCGTGTGAGCCACCATCAATACAGAAAGTACAGATGTA 2589
Db 2521 TCAGAGGGCCCCCACTCCGTGTGAGCCACCATCAATACAGAAAGTACAGATGTA 2580
OY 2590 CCAGCAAGATGCATCCCTCTTTTGTGTGGGGGTATGAGCTCCAAAAGCCAGCTGAA 2649
Db 2581 CCAGCAAGATGCATCCCTCTTTTGTGTGGGGGTATGAGCTCCAAAAGCCAGCTGAA 2640
OY 2650 CAATTAATAATGATTGAGC 2669
Db 2641 CAATTAATAATGATTGAGC 2660

RESULT 3
AF320294 2022 bp mRNA 1linear PRT 13-DEC-2000
LOCUS Homo sapiens ABCG8 (ABCG8) mRNA, complete cds.
DEFINITION AF320294
ACCESSION AF320294.1 GI:11692801
VERSION AF320294.1
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 2022)
Berge, K.E., Tian, H., Graf, G.A., Yu, L., Grishin, N.V., Schultz, J.,
Kwiterovich, P., Shan, B., Barnes, R. and Hobbs, H.H.,
Accumulation of Dietary Cholesterol in Siltostolemia Caused by
Mutations in Adjacent ABC Transporters

JOURNAL Science (2001) In press
2 (bases 1 to 2022)
Berge, K.E., Tian, H., Graf, G.A., Yu, L., Grishin, N.V., Schultz, J.,
Kwiterovich, P., Shan, B., Barnes, R. and Hobbs, H.H.,
Direct Submission

TITLE Submitted (09-NOV-2000) Molecular Genetics, University of Texas,
Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd.,
Dallas, TX 75390-9046, USA

FEATURES
source Location/Qualifiers

1..2022
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="Caxon:9606"
1..2022
/gene="ABCG8"
1..2022
/note="ATP-binding cassette, subfamily G, member 8"
/codon_start=1
/product="ABCG8"
/protein_id="AAG40004.1"
/db_xref="GI:11692802"

/translation="MAGKAEREGIPKATPDITSDGLDRLPSESSEDSLYFTYSGP
MTLEVRDLATQVDAISQVPMFEQLAQFMPTSPSCQNSCELGIONLSFKYRSGMLA
ITGSSGCGRASLDIVITGSGHGKIKSQIWIINGPSPQVLRKCAVATROHNLIPN
LTVRETLAFIAOMRLPTFSOARDKREVDYIAELRLSCADTRGVNAYVGLSGER
RRLVIGVOLLNPGILIDEPSTGSLDSTANLVKTSRLAKGNRLVYISLHORSDDI
FRFLDLVLTSGPIYGAOHMYOFTATGTCPRKSNADPYVDITSIDRSRBO
ELATREKRSALAFIEKVRDLDPFKAEKRDLEDVCESVTPDITNCLSPSTKM
PGAVQATTLIRROISNDPRLPTLLHGAACLSMTIGELVGHGSIQSPQATA
LFTMGALIPNVILVDVYSKYSERAMLYLEDEGLYTGPFYFAKILGELPEACAYI
IYVPPYTMALNRPGLPLHLFLVLVVFCCRIALAAALPFPFMSFNAL
INSEYLAGGEMINLSLMTVPMASIKVSFLNCFEGGLKIOFSRRTYMPMLGNLTIAV
SGDKILSYNELDSPYLAIVLYIVIGLSGFEVLYVYSIRFKORSDM"

BASE COUNT 424 a 603 c 545 g 450 t
ORIGIN

Query Match 75.7%; Score 2020.4; DB 9; Length 2022;
Best Local Similarity 100.0%; Pred. No. 0;

Matches 2021: Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 100 ATGGCCGGGAAGCGCGAGAGAGAGAGAGGCTCCGAAAAGGGCCACTCCAGAGATAC 159
Db 1 ATGGCCGGGAAGCGCGAGAGAGAGAGAGGCTCCGAAAAGGGCCACTCCAGAGATAC 60
OY 160 TCGGGGCTCCAGATAGATTTGTTCTCTCTGAAAGTGACAAAGCTGTACTTCACTTAC 219
Db 61 TCGGGGCTCCAGATAGATTTGTTCTCTCTGAAAGTGACAAAGCTGTACTTCACTTAC 120
OY 220 AGTGCCAGGCCCAACACCCCTGGAGGTGAGAGACTCACTACAGGTGAGACCTGGGCTCT 279
Db 121 AGTGCCAGGCCCAACACCCCTGGAGGTGAGAGACTCACTACAGGTGAGACCTGGGCTCT 180
OY 280 CAGGTCCTTGTGTTGAGCAGCTGCTCAGTTCAAGATGCCCTGACATCTCCAGCTGC 339
Db 181 CAGGTCCTTGTGTTGAGCAGCTGCTCAGTTCAAGATGCCCTGACATCTCCAGCTGC 240
OY 340 CAGATTTCTTTGAGCTGGGCAATCCAGAACCTAAGCTTCAAGTGAAGTGAAGTGGCAGATG 399
Db 241 CAGATTTCTTTGAGCTGGGCAATCCAGAACCTAAGCTTCAAGTGAAGTGAAGTGGCAGATG 300
OY 400 CTGGCCATCATAGGAGGAGCTCAGGTTGGGAGAGCCCTCTTGTAGATGTGATCATCTGC 459
Db 301 CTGGCCATCATAGGAGGAGCTCAGGTTGGGAGAGCCCTCTTGTAGATGTGATCATCTGC 360
OY 460 CGAGGTACAGGGCGGCAAGATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 519
Db 361 CGAGGTACAGGGCGGCAAGATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 420
OY 520 CCTCAGCTGTGAGAGAGTGTGTGGCCACGTGCGCCACAGACAGCTGCTCCCAAC 579
Db 421 CCTCAGCTGTGAGAGAGTGTGTGGCCACGTGCGCCACAGACAGCTGCTCCCAAC 480
OY 580 TTGACGTGTGAGAGAGCTTGGCCCTTCAATGGCCAGATGCGGCTGCCAGAACCTTCTCC 639
Db 481 TTGACGTGTGAGAGAGCTTGGCCCTTCAATGGCCAGATGCGGCTGCCAGAACCTTCTCC 540
OY 640 CAGGCCCAAGCTGTACAAAAGAGGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGC 699
Db 541 CAGGCCCAAGCTGTACAAAAGAGGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGC 600
OY 700 GCTGACACCCGCTGTGGCAACATGTACGTGCGGGGTTGTGCGGGGGTGTGAGCGCAGAGA 759
Db 601 GCTGACACCCGCTGTGGCAACATGTACGTGCGGGGTTGTGCGGGGGTGTGAGCGCAGAGA 660
OY 760 GTGACATTTGGGGGTGAGCTCTCTGTGGAACCCAGAGATCTTATCTGTGAGCAACCCAC 819
Db 661 GTGACATTTGGGGGTGAGCTCTCTGTGGAACCCAGAGATCTTATCTGTGAGCAACCCAC 720
OY 820 TCTGGGCTGAGAGCTTACAGCCCAACCTGTGAGAGACCTTGTCCAGGCTGGCCAAA 879
Db 721 TCTGGGCTGAGAGCTTACAGCCCAACCTGTGAGAGACCTTGTCCAGGCTGGCCAAA 780
OY 880 GCGAACCGGCTGTGCTCATCTCTCTCCACAGAGCTGCTGTGACATCTTCAAGCTGTTT 939
Db 781 GCGAACCGGCTGTGCTCATCTCTCTCCACAGAGCTGCTGTGACATCTTCAAGCTGTTT 840
OY 940 GATCGGTCCTCCGATATGAGCTGTGAGACCCCACTACTTATGAGGGCGGCCAGACATG 999
Db 841 GATCGGTCCTCCGATATGAGCTGTGAGACCCCACTACTTATGAGGGCGGCCAGACATG 900
OY 1000 GTCCAGTATTTACAGCAGCATGCGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1059
Db 901 GTCCAGTATTTACAGCAGCATGCGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
OY 1060 TATGTGAGCTGAGCAGCATGAGAGGCGCAGAGAGAGAGAGATTTGGCCACAGGAG 1119
Db 961 TATGTGAGCTGAGCAGCATGAGAGGCGCAGAGAGAGAGAGATTTGGCCACAGGAG 1020
OY 1120 AAGGCTCAGTACCTGCGAGCCCTGTTTCTAAGAAAAGGCTGACTTGAATGACTTTCTA 1179
Db 1021 AAGGCTCAGTACCTGCGAGCCCTGTTTCTAAGAAAAGGCTGACTTGAATGACTTTCTA 1080

SOURCE ORGANISM	Hom sapiens (human)
REFERENCE	Hom sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1
JOURNAL	Tang, Y.-T., Yue, H., Nguyen, D.B., Hafalila, A.J., Elliott, V.S., Lu, Y., Walla, N.K., Yao, M.G., Baughn, M.R., Gandhi, A.R., Ding, L., Sanjanwalala, M., Ramkumar, J., Atvizu, C., Gletzen, K.D., Lal, P.G., Azimzai, Y., Khan, F.A., Thangavelu, K., Thornton, M., Lu, D.A., Tridbouley, C.M., Warren, B.A., Ison, C.H., Das, D., Raumann, B.E., Policky, J.L. and Kearney, L. Transporters and Ion channels Patent: WO 0240541-A 29-23-MAY-2002; Incyte Genomics, Inc. (US)
FEATURES	Location/Qualifiers
SOURCE	1..3239
BASE COUNT	784 a 822 c 796 g 837 t
ORIGIN	784 a 822 c 796 g 837 t
Query Match	63.0%; Score 1680.6; DB 6; Length 3239;
Best Local Similarity	99.8%; Pred No. 0;
Matches 1683;	Conservative 0; Mismatches 4; Indels 0; Gaps 0;
983	GGGGGCCAGACATGTCAGATATTTACAGAGCCATCGGTACCCCTGCTCGCTACA 1047
12	GGGGGGGGCCACACATGTCATATTTACAGAGCCATCGGTACCCCTGCTCGCTACA 71
1043	GCAATCCGTCGATCTTATGTCGACCTCGACAGATTTACAGAGGGCGACAGAGCAGG 1107
72	GCAATCCGTCGATCTTATGTCGACCTCGACAGATTTACAGAGGGCGACAGAGCAGG 131
1103	AATGGCCACACAGGAGAGGCTCAGTACTGCGACCCCTGTTCTAATAAAGTGGCG 1167
132	AATGGCCACACAGGAGAGGCTCAGTACTGCGACCCCTGTTCTAATAAAGTGGCG 191
1153	ACTTATGATGATCTTCTATGGAAGACAGACAGAGAGATCTTGACAGAGACCTGTGG 1222
192	ACTTATGATGATCTTCTATGGAAGAGAGAGAGATCTTGACAGAGACCTGTGG 251
1223	AAAGCAGAGTACCCCATATGACACCAACTCGTCGAGTCTACGAAAGATGCTGGGG 1287
252	AAAGCAGAGTACCCCATATGACACCAACTCGTCGAGTCTACGAAAGATGCTGGGG 311
1283	CGGTGACAGATTACGAGCGCTGATCCGTCGTCAGATTTCCAGACACTTCGAGACTGC 1347
312	CGGTGACAGATTACGAGCGCTGATCCGTCGTCAGATTTCCAGACACTTCGAGACTGC 371
1343	CCACCCCTGTCATCCATGAGGGGGGAGGCGCTGTCATGATGATGATGATGATGATGAT 1407
372	CCACCCCTGTCATCCATGAGGGGGGAGGCGCTGTCATGATGATGATGATGATGATGAT 431
1403	ATTTTGGGCATGGAGCATCCAGCTGTCCTTCATGATGATGATGATGATGATGATGATGAT 1467
432	ATTTTGGGCATGGAGCATCCAGCTGTCCTTCATGATGATGATGATGATGATGATGATGAT 491
1463	TCGGTGCCTCATCCCTTTCAAGCTCATTTGATGATGATGATGATGATGATGATGATGATGAT 1522
492	TCGGTGCCTCATCCCTTTCAAGCTCATTTGATGATGATGATGATGATGATGATGATGATGAT 551
1523	GGGCAATGCTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1582
552	GGGCAATGCTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 611
1583	CCAAGATCTTCGGGAGCTTCGGAGACAGTGTGCTATCATCATCATCATCATCATCATCATCAT 1642
612	CCAAGATCTTCGGGAGCTTCGGAGACAGTGTGCTATCATCATCATCATCATCATCATCATCAT 671
1643	CTCTGTCGGGCGCAACTGAGGGCAGGCGCCCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1702

[illegible]

[illegible]

OY 936 GTTATGCTGCTCTGATGACGTCTGGACACCCCACTTACTTAGGGGGGCCACGA 995
 DB 941 ATTTAACTGCTCTTGTATGACATCTGGACACCCCTTACTTACCTGGGGGGCCACCA 1000
 OY 996 CATGCTCAGATTTCACAGCAGTGGGCTACCCCTGCTCTGCTACCAATTCCTGCTGA 1055
 DB 1001 ATGCTGACGACTTACATTCATTCATGGCCACCTTGTCTGCTATACCAACCTGGCGGA 1060
 OY 1056 CTCTATGTGACCTGACACAGATTGACAGGGGCGACAGAGAGAGAGAAATGGCCACAG 1115
 DB 1061 CTCTACGCTGACCTGACAGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1120
 OY 1116 GGAGAGGCTCAGTACTGACAGCCCTTCTTCTGAAAAAGTGGCTACTAGATGACTT 1175
 DB 1121 GGAGAGGCTCAGTACTGACAGCCCTTCTTCTGAAAAAGTACAAAGCTTGTATGACTT 1180
 OY 1176 TCTATGAAAG 1235
 DB 1181 TCTGAGAAAGCTGAG 1240
 OY 1236 CCCACTAGACACCACTGCTCCGAGTCTACAGAGATGCTGGGGGGCTGACAGACTT 1295
 DB 1241 CACACAGACACTGACTG---TGAGACTGCTGTGAGCTGCGCGGAGATGAGAGACTT 1297
 OY 1296 TACGAGCTGATCCGCTGCTCAGATTTCACAGACTTCCGAGACTGCGCCACCTCTCAT 1355
 DB 1298 TTCCACCTGATCCGCTGCTCAGATTTCACAGACTTCCGAGACTGCGCCACCTGCTCAT 1357
 OY 1356 CCATGGGGGAGAGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1415
 DB 1358 TCAATGGGCTGAGAGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1417
 OY 1416 GAGACCTCAGCTCTCTCTCATGATGATGATGATGATGATGATGATGATGATGATGATG 1475
 DB 1418 GGCACAGAGCTCTCTCTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1477
 OY 1476 CCCCTTCAAGCTCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1535
 DB 1478 TCTCTTCAATGCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1537
 OY 1536 CTATGAACTGAGAGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1595
 DB 1538 CTATGAACTGAGAGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1597
 OY 1596 GGAGCTTCCGAGAGCTGCTCATCATCATCATCATCATCATCATCATCATCATCATCAT 1655
 DB 1598 AGAATTCGCGAGAGCTGCTCATCATCATCATCATCATCATCATCATCATCATCATCAT 1657
 OY 1656 CAACCTGAGGAGAGGCTTCCAGCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1715
 DB 1658 AACCCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1717
 OY 1716 CTCTGTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1775
 DB 1718 CTCTGTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1777
 OY 1776 CTCTCTTCAAGAGAGGCTTCTCAACTCTTCTACCTGCGCGGGCTTCATGATGATAA 1835
 DB 1778 CTCTCTTCAAGAGAGGCTTCTCAACTCTTCTACCTGCGCGGGCTTCATGATGATAA 1837
 OY 1836 CTGAGAGAGCTGAG 1895
 DB 1838 CTGAGAGAGCTGAG 1897
 OY 1896 TTTTGAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1955
 DB 1898 CTCTCGGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1957
 OY 1956 CTCTACCATGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2015
 DB 1958 CTCTACCATGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2017
 OY 2016 TCTCTAGGCACTACTCATCTCATCTCATCTCATCTCATCTCATCTCATCTCATCTA 2075

DB 2018 ACTATATCGATCTACTCATATGATGATGATGATGATGATGATGATGATGATGATGATG 2077
 OY 2076 CGTGTCTTAAGTCTATCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2135
 DB 2078 TCTATCTTGAAGCTATCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2137
 OY 2136 CTGCGCGCTGGTGGGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2195
 DB 2138 CTCACTGGCG-----GACCTTTTCCGGGGCTGGCCACCCAGAGAGAGAGAGAGAG 2185
 OY 2196 TTCTGGGGAG 2255
 DB 2186 GGAGCTGGGGAG 2238
 OY 2256 CTGAGAGGAG 2315
 DB 2259 CTGAGAGGAG 2298
 OY 2316 CTGCTACTGAG 2375
 DB 2299 CTGCTCCAG 2343
 OY 2376 AGTGTCTAATTTATTTCTTTTATGATGATGATGATGATGATGATGATGATGATGATG 2435
 DB 2344 AGTGTCTAATTTATTTCTTTTATGATGATGATGATGATGATGATGATGATGATGATG 2395
 OY 2436 AGCAATCAG 2495
 DB 2396 AACCATTATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2455
 OY 2496 ACAATTAACAG 2555
 DB 2456 ATAGTGTATATGAG 2511
 OY 2556 GCCAATCAG 2614
 DB 2512 GCCAATCAG 2571
 OY 2615 TGTGTGGGCTATGAG 2646
 DB 2572 TGTGTGGGCTATGAG 2603

RESULT 7
 AY196216
 LOCUS
 DEFINITION
 Mus musculus strain PERA/EI ATP-binding cassette sub-family G
 member 8 (Abcg8) mRNA, complete cds.
 ACCESSION
 AY196216
 VERSION
 AY196216.1 GI:31322261
 KEYWORDS
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 2284)
 Wittenburg, H., Lyons, M.A., Li, R., Churchill, G.A., Carey, M.C. and
 Paigen, B.
 Primary Roles of FXR and ABCG5/ABCG8 in cholesterol Gallstone
 Susceptibility: Evidence from a Cross of PERA/EI and I/Jn Inbred
 Mice
 Unpublished
 2 (bases 1 to 2284)
 Lyons, M.A., Wittenburg, H., Walsh, K.A., Carey, M.C. and Paigen, B.
 Direct Submission
 Submitted (12-DEC-2002) The Jackson Laboratory, 600 Main Street,
 Bar Harbor, ME 04609, USA
 Location/Qualifiers
 1..2284
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="PERA/EI"

```
/db_xref="taxon:10090"
/chromosome="17"
/map="55 cm"
/sex="male"
/tissue_type="liver"
1..2284
/gene="Abcg8"
102..2120
/gene="Abcg8"
/feature="ATP-dependent canalicular cholesterol transporter;
white subfamily"
/codon_start=1
/product="ATP-binding cassette sub-family G member 8"
/protein_id="AA045096.1"
/db_xref="GI:3132262"
/translation="MAEKREYQVLMNGVLDASGLDLSFSSSDNSLYFTYSGOS
NLEVDLTYQVDIAQVWPEQLAOFKIPWRSHSSQSDCELRIMLSEFRSGOMLA
11GSSGCGRASLDVITGRGHGKMSGOIMINGOSPQVLYKCAVYAROHOLLPN
LVRETLIAQMRLEPFSOARDKRDVLAELRLROCANTRVGNVYGVSGGR
RVSIGVOLMNGVILIDEPSTGLSPFAHNLVTLISLANRNLVILISINPRSDI
PRLFDVILMTSGTPYVGAQAOQVYFTSIGHPCRYNRPADFTVLDLSIDRSER
EVATYERKASIALALEKYQGFDDFLMKAELAKLNTSTHVSLLTLDYDQGVLEP
GMIBOSTLIRKQISNDRDLPLLIHGEACMLSLIGFLYGHGAKOLSEFDATL
LEMIGALIPFNVLIDVSKHSRMLYELLEDGLYTAGYFEAKLIGELPEHCAYI
IYAMPYVNLNRPVLEFLHFLHFLVWLVFCORMALASAMLPENHSPFCNALY
NSFYLAGEFMINDLMIVPAMISKLSFLRMCESGIMOIOFNHLYTTQIGNTTSIL
GPMISAMDLNHPYAIYLIYIGISYGLFLYISLKLKIKSIDW"
BASE COUNT      514 a      676 c      591 g      503 t
ORIGIN
Query Match      54.5%; Score 1454.2; DB 10; Length 2284;
Best local Similarity 79.8%; Pred. No. 2,3e-311;
Matches 1760; Conservative 0; Mismatches 423; Indels 22; Gaps 3;
QY 99 CATGCCGGGAAGGCGGAGAGAGAGAGAGGCTGCCGAAAGGGCCACTCCCGAGATAC 158
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 101 CATGCTGAGAAACCAAGAAAGAGACCAGCTGTGAATGGGAGCTGTACTTCAGATAC 160
QY 159 CTCGGGCTCCAGAGATAGTTGTCTCCCTGAAAGTGCACACAGCTTACTTACCTA 218
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 161 TTGGGGCTCCAGAGAGAGCTGTCTCTCCGGAAGTGCACACAGCTTACTTACCTA 220
QY 219 CAGTGGCCAGCCCAACACCTCGAGGTGAGAGACCTCACTACGAGGTGAGACCTG 278
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 221 CAGTGGTACGTCCACACCTCTGAGGTGAGATCTCACCTACAGGTGAGATGCTCTC 280
QY 279 TCAGTCCCTTGGTTGAGACAGCTGCTCAGTTCAAGATGCCGTGACATCTCCACGT 338
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 281 TCAGTGGCTTGTGAGACAGCTGCTCAGTTCAAGATACCTGAGGTGACATGACAG 340
QY 339 CGAAGATCTTGTGAGCTGGGATCCAGAACCTTAAGCTTCAAGTGAAGTGGGCAGAT 398
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 341 CCAAGACTCTGTGAGCTGGGATCCGAATCTTAAGCTTCAAGTGAAGTGGGCAGAT 400
QY 399 GCTGGCCATCATAGGAGCTCAGTTGTGGGAGAGCCCTCTGTCAGATGTCATACATG 458
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 401 GCTGGCCATCATAGGAGCTCAGCTGCGGGAGAGCCCTCCTACTACGACGTATACAG 460
QY 459 CCGAGGTACGGGCGCAAGATCAAGTCAAGCCAGATCTGATCAATGGGAGAGCTG 518
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 461 CAGAGGCCACGCTGCGCAAGATCAAGTCAAGCAAAATTTGAAATTTGGGCAACCCAG 520
QY 519 GCTTCAGTGTGAGAGAGTGTGTGCGCCAGCTGCGCCAGCAACAAGCTGCTCCAA 578
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 521 GCTTCAGTGTGAGAGAGTGTGTGCGCCAGCTGCGCCAGCAACAAGCTGCTCCAA 580
QY 579 CTTGACTGTGAGAGAGCTTGGCTTCAATGGCCAGATGGGCTGCGCCAGCAACAAG 638
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 581 CTTGACTGTGAGAGAGAGCTTGGCTTCAATGGCCAGATGGGCTGCGCCAGCAACAAG 640
QY 639 CCAGGCCACAGGCTGCAAAAAGGTGAGAGAGCTGTCGAGAGCTGCGGCTTACGAG 698
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 641 CCAGGCCACAGGCTGCAAAAAGGTGAGAGAGCTTACGCCAGAGCTGCGGCTGCGCAAG 700
```

```
QY 699 CGCTGACACCCCGCTGGGCAACATGTACGTGCGGGGTTGTGCGGGGTGAGCGCAGAG 758
    ||| ||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 701 CGCCACACACAGAGGTGGGCAACAGATGTACGTGCGGGTGTCCGGGGTGAGCGCGAG 760
QY 759 AGTCAGATTTGGGGTGCAGCTCCGTCGGAACCCAGGAATCTTATCTGCACAGCCAC 818
    ||| ||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 761 AGTCAGATTTGGGGTGCAGCTCCGTCGGAACCCAGGAATCTTATCTGCATACCCAC 820
QY 819 CTCTGGGCTGCAGAGCTTCCAGCCACCAACCTGTGTGAAGACCTTGTCCAGGTGG 878
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 821 TTCTGGGCTGCAGAGCTTCCAGCCACCAATCTGTGTACACCTTGTCCCGCTGAG 880
QY 879 AGCCACACGCTGTGCTCATCTCCACACAGCTGCTGACATCTTACAGCTGT 938
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 881 GGGCAACAGCTGTGCTCATCTCCACACAGCTGCTGACATCTTACAGCTGT 940
QY 939 TGATCTGCTCTCTGATGACGCTGTGACACCCCATCTTACTTATGGGGCGCCAG 998
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 941 TGACCTGCTCTCTGATGACATCTGACACCCCATCTTACTTATGGGGCGCCAG 1000
QY 999 GGTCCAGTATTTACAGACCTACGCTACCCCTGCTCCGCTACAGAAATCCGCTACT 1058
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1001 GTTGCACTACTTACATCTTGTGCAACCTTGTCTCGCTATAGCAACCTGCGGACT 1060
QY 1059 CTATGTGACCTGACAGCATTTGACAGGGGCGACAGAGACAGGAATGGCCACAG 1118
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1061 CTACGTGACCTTACAGCATTTGACAGGGGCGACAGCAAGAGAGAGTGGCCAC 1120
QY 1119 GAAGCTCAGTACTGCGACCCCTGTTTACAAAAAGTGTGATGATGATCTTCT 1178
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1121 GAAGGACAGATCTTGTGACAGCCCTGTTCTAGAAAAAGTGAAGCTTGAATGAT 1180
QY 1179 ATGAAGAAGAGAGACAGAGATTTGACAGGACACCTGTGTGGAAGAGAGCTGAC 1238
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1181 GTGGAAGCTGTGAGGCAAGAGACTCAACACAGACACACAGTACAGCTGACCT 1240
QY 1239 ACTAGACACCAACTGCTCCGAGTCTACGAAGATGCTGGGGCGGTGACAGATTTAC 1298
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1241 ACAGGACACTGACTG--TGGGAGCTGTGTAGCTGCGCGGAGATGATGACAGAT 1297
QY 1299 GACGCTGATCCGTCGATATTTCCACAGACTTCCGAGACTGCCACCTCTCATCCA 1358
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1298 GACCTGATCCGTCGATATTTCCACAGACTTCCGAGACTGCCACCTCTCATCCA 1357
QY 1359 TGGGGGAGAGGCTGTGATGTCATAGCAATGCAATGCGCTCTCTATTTTGGGCA 1418
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1358 TGGGTCGGAAGCTGCTGATGTCCTCTCATATGCTGCTCTTACTACGGGCAAT 1417
QY 1419 CATCCAGCTCTCTTCATGATGATGATGACAGCCGCTCTTGTTCATGATGCTCTAT 1478
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1418 CAAGGAGCTCTCTTCATGATGATGATGACAGAGCCCTCTCTTCATGATGATG 1477
QY 1479 TTTCAAGCTGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1538
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1478 TTTCAAGCTGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1537
QY 1539 TGAAGTGAAGAGGAGCTGTACAGCAAGTGTGATGATGATGATGATGATGATGAT 1598
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1538 TGAAGTGAAGAGGAGCTGTACAGCAAGTGTGATGATGATGATGATGATGATGAT 1597
QY 1599 GCTTCGAGAGCTGCTTACATATCATATCATATCATATCATATCATATCATATCAT 1658
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1598 ATTCGCGAGAGCTGTGCTTACATATCATATCATATCATATCATATCATATCATAT 1657
QY 1659 CTTGAGGCGAGGCTTCCAGCCCTCTCTGATGATGATGATGATGATGATGATGAT 1718
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1658 CTTGAGGCGAGGCTTCCAGCCCTCTCTGATGATGATGATGATGATGATGATGAT 1717
QY 1719 CTTGAGGAGATTAAGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1778
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1718 CTTGAGGAGATTAAGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1777
```


OY 1779 CTTCTTACGAAATGCGCTCTACACTCTTACCTCGCGGGGGCTTCATGATAACTT 1838
 DB 1778 CTTCTTACGAAATGCGCTCTACACTCTTACCTCGCGGGGGCTTCATGATAACTT 1837
 OY 1839 GAGCAGCTGTGGAGAGTCCCGCGTGAATTCGAAAGTCTCTCGGGTGTGTTT 1898
 DB 1838 GAGCAGCTGTGGAGAGTCCCGCGTGAATTCGAAAGTCTCTCGGGTGTGTTT 1897
 OY 1899 TGAAGGCTGATGAAGATTCAGTTCAGCAGAAAGAACTTAAATGCTCTCGGAACTT 1958
 DB 1898 CTGGGGGCTGATGAGATTCATTTAAATGACCTTTACACCAACAAATCGGCAACTT 1957
 OY 1959 CACCAATCGCGGTCTAGAGATTAATTCCTAGTGCAGTGGAGCTGAGCTGACCTTCT 2018
 DB 1958 CACCTTCTGCAATCTCGGAGACATGATGATGCTGACCTGAGCTGAGCTGACCTTCT 2017
 OY 2019 CTAGCCCATCTACCTCATCTGCTATGCTGCTGAGCTGCTGCTGCTGCTGCTGCTGCT 2078
 DB 2018 CTATCGATCTACCTCATCTGCTGCTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCT 2077
 OY 2079 GTCTTAAGTTCATCAACAGAAACCAAGTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 2138
 DB 2078 ATCTTGAAGCTCAACAGAAACCAAGTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2137
 OY 2139 CCCCCTGTGGGGGAGCTGAGAGACCTTCAACTGCACTCCCTGCTGAGAGCCCTTCT 2198
 DB 2138 ACTGGCGG-----GACCTTTTCCCGGGGCTGGCCACCCAGAGAGCCCGGA 2185
 OY 2199 CTGGGGAGACGTGAGACATGACCTTACAGATGCTCAGCTACATCCGGCCAGGGTCTG 2258
 DB 2186 CTGGGGAGACGTGAGACATGACCTTACAGATGCTCAG-----GCAGCAGCACCCTCTTAATGCTG 2238
 OY 2239 CAGTGGCAGACACGACCAAGAGATGCAAGTGAATTAAGACAGT 2303
 DB 2239 CAGTGGCAGACGCTGAGCCACAGAGATGCAAGTGAATTAAGACAGT 2283
 RESULT 8
 AY196215 2285 bp mRNA linear ROD 01-JUN-2003
 LOCUS Mus musculus strain I/LmJ ATP-binding cassette sub-family G member
 DEFINITION 8 (Abcg8) mRNA, complete cds.
 ACCESSION AY196215
 VERSION AY196215.1 GI:31322259
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Wittenburg, H., Lyons, M.A., Li, R., Churchill, G.A., Carey, M.C. and
 Paigen, B.
 TITLE Primary Roles of FXR and ABCG5/ABCG8 in Cholesterol Gallstone
 Susceptibility: Evidence from a Cross of PERA/EI and I/Lm Inbred
 Mice
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2285)
 AUTHORS Lyons, M.A., Wittenburg, H., Walsh, K.A., Carey, M.C. and Paigen, B.
 TITLE Direct Submission
 JOURNAL Submitted (12-DEC-2002) The Jackson Laboratory, 600 Main Street,
 Bar Harbor, ME 04609, USA
 FEATURES
 source Location/Qualifiers
 1..2285
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="I/LmJ"
 /db_xref="taxon:10090"
 /chromosome="17"
 /map="55 cM"
 /sex="male"
 /tissue_type="liver"
 1..2285
 /gene="Abcg8"

CDS
 102..2120
 /gene="Abcg8"
 /note="ATP-dependent canalicular cholesterol transporter;
 white subfamily"
 /codon_start=1
 /product="ATP-binding cassette sub-family G member 8"
 /protein_id="AA045095.1"
 /db_xref="GI:31322260"
 /translation="MAEKTEQYOLANGLVLDASGLDSESSDSNLSVFTYSGOS
 NLEFRLDYVDIASOYPMFEOLOAFIPMHSOSDCEGLINLSFKYRSGOMLA
 IIGSSGGRASLDVITRGHGRKRSQIYNQPSPOLYRKRVAVHYRHDILPN
 LIVERETLAFIDMRIPREFSOQRKRVDEYIAELRLQCAKTRVGNVYGVSGER
 RNVISGVLLNMPGILILDEPTSGDSTFANHLVTLRLAKNRLVLSLSHQPSDI
 FLEFDVLMTSGPIYLGADQMOVQYTSIGHPCPRYSNPADFPVDTLSIDRSKER
 EVATVEKQSLAALELEKVOGDEPDLAEKELSTHYSLTLTODTDCGTAELEP
 GMIEQFSTLIRROISNDPRDLPTLLIHSEACLSMLIIGPLYYGAGKOLSPMDTAL
 LEMIALPFWITLIDVYKRSERMSLVELEDGLYTGPPFPAKILGEALDEPHCAVI
 IYAMPITYLNLRPPELFLHLHLVYLVRCCRTMALAASMLPTFMSSFFCALY
 NSFYLITAFMINDLMTVPAMISLNSLRFCSLMDIQRNGHLYTQIINFPSIL
 GPTMISADNLNHPDYAIYLVIGISYGFELYYLSLKLIRKSIQDW"
 BASE COUNT 518 a 679 c 591 g 497 t
 ORIGIN
 Query Match 54.3%; Score 1449.4; DB 10; Length 2285;
 Best Local Similarity 79.8%; Pred. No. 2.6e-310;
 Matches 1755; Conservative 0; Mismatches 421; Indels 23; Gaps 3;
 OY 99 CATGGCCGGAGGAGGGGAGAGAGAGGGGTGCCGAAAGGGCCACTCCCAAGAGATAC 158
 DB 101 CATGGCTGAGAAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 160
 OY 159 CTGGGGCCCTCCAGATAGATTGTTCTCTCTCAAGATGACACAGCTGTACTTACCTA 218
 DB 161 TTGGGGCCCTCCAG 220
 OY 219 CAGTGGCCACCCCAACACCTTGGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 278
 DB 221 CAGCGGTACGTCACCAACACTCTGAGGTGAGAGATCTACCTACAGGTGAGAGAGAG 280
 OY 279 TCAGTCCCTGTTGAG 338
 DB 281 TCAGTCCCTGTTGAG 340
 OY 339 CCAGATTCCTGTGAGTGGGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 398
 DB 341 CCAAGACTCCTGTGAGTGGGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 400
 OY 399 GCTGGCCATCATAGGAG 458
 DB 401 GCTGGCCATCATAGGAG 460
 OY 459 CCGAGTCAAGGAGGAG 518
 DB 461 CAGAGGCAACGAGTGGCAAGATGAAATGACGAAATTTGGATTAACGGGCAACCAAGTAC 520
 OY 519 GCGTCAAGTGTGAG 578
 DB 521 GCGTCAAGTGTGAG 580
 OY 579 CTGACTGTGGAG 638
 DB 581 CCGTCAAGTGTGAG 640
 OY 639 CAGGAGCCAGGCTGACAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 698
 DB 641 CAGGAGCCAGGCTGACAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 700
 OY 699 CGCTGACACCCGCTGGGCAACATGTACGTGCGGGGGTGTGCGGGGGTGAAGCCAGAG 758
 DB 701 CGCCAAACACAGATGGGCAACAGATGTACGTGGGGGTGCGGGGGTGAAGCCAGAG 760
 OY 759 AGTCAGATTGGGGTGCAGCTCCCTGTGAAGACCAAGAACTTATTCTCAACGAAACCAC 818

Db 761 AGTACGATTTGGGTGACGCTCCGTGGAACCCAGAAATCTCATTCGTGATGAACCCAC 820
QY 819 CTCGGGCTGGACAGCTTCACAGCCCAACCTGGTGAAGACCTTGTCCAGGCTGGCAA 878
Db 821 TTCTGGCTTGGACAGCTTCACAGCCCAACCTGGTGAAGACCTTGTCCAGGCTGGCAA 880
QY 879 AGGCAACCGGCTGGTGTCTCTCCCTCCAGCCCTGGTGTGATATCTTCAGGCTGT 938
Db 881 GGGCAACAGGCTGGTGTCTCTCCCTCCAGCCCTGGTGTGATATCTTCAGGCTGT 940
QY 939 TGAATGTGCTCTGTATGACAGCTGGACCCCATCTTACTTAAAGGCGGCGGACGACAT 998
Db 941 TGACCTGTCTCTGTATGACAGCTGGACCCCATCTTACTTAAAGGCGGCGGACGACAAAT 1000
QY 999 GGTGACGATTTATGACAGCCATGGGCTACCCCTCTCTGCTACAGCAATCTGTGACTT 1058
Db 1001 GGTGACGATTTATGACAGCCATGGGCTACCCCTCTCTGCTACAGCAATCTGTGACTT 1060
QY 1059 CTATGTGACCTGACAGCATTTGACAGGCGCAGACAGAGCAAGAAATGGCCACAGGGA 1118
Db 1061 CTATGTGACCTGACAGCATTTGACAGGCGCAGACAGAGCAAGAAATGGCCACAGGGA 1120
QY 1119 GAAAGCTCAGTCTGACAGCCCTGTTTCTAGAAAAATGCTGACTTAAATGACTTCT 1178
Db 1121 GAAAGCTCAGTCTGACAGCCCTGTTTCTAGAAAAATGCTGACTTAAATGACTTCT 1180
QY 1179 ATGGAAGCAGACAGCAAGATTTTACAGAGACACCTGTGTGAAAAAGCAGCTGACCC 1238
Db 1181 GTGGAAGCTGAGCAAGAAATGCTTCAACACACACACAGCTGACCTGACCTGAC 1240
QY 1239 ACTAGACACCAACTGCTCCGAGTCTCTAGAAAGATGCTGGGGGCGTGCAGCAGTTAC 1298
Db 1241 ACAGGACACTGACTG---TGGACACTCTCTGACTGCTCCCGGATGATAGAGAGTTTC 1297
QY 1299 GACGCTGATCCGCTGTCAGATTTTCAACAGCTTCGAGACCTGCCACCTCTCATTTCA 1358
Db 1298 CACCTGATCCGCTGTCAGATTTTCAACAGCTTCGAGACCTGCCACCTCTCATTTCA 1357
QY 1359 TGGGGGGAGGCTGTCTGATGTCATGACATGACATGCGCTCTCTATTTTGGCCAGGAG 1418
Db 1358 TGGGGGGAGGCTGTCTGATGTCATGACATGACATGCGCTCTCTATTTTGGCCAGGAG 1417
QY 1419 CATCCAGCTCTCTCTGATGATGATACAGCCGCTCTGATGATGATGATGATGATGATGAT 1478
Db 1418 CAGAGGCTCTCTCTGATGATGATACAGCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1477
QY 1479 TTTCAAGCTCATTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1538
Db 1478 TTTCAAGCTCATTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1537
QY 1539 TGAAGCTGGAAGAGGGCTGTACACAGCTGGTCCATATTTCTTTGCCAAGATCTCGGGGA 1598
Db 1538 TGAAGCTGGAAGAGGGCTGTACACAGCTGGTCCATATTTCTTTGCCAAGATCTCGGGGA 1597
QY 1599 GCTTCGAGAGCACTGTGCTTACATCATCTTACAGGAGGAGGAGGAGGAGGAGGAGGAGG 1658
Db 1598 ATTTGCGGAGCACTGTGCTTACATCATCTTACAGGAGGAGGAGGAGGAGGAGGAGGAGG 1657
QY 1659 CCTGAGGAGGAGGCTTCACAGCCCTCTCTGCTGACTTCTCTGCTGCTGCTGCTGCTGCT 1718
Db 1658 CCTGAGGAGGAGGCTTCACAGCCCTCTCTGCTGACTTCTCTGCTGCTGCTGCTGCTGCT 1717
QY 1719 CTGTTGCAAGATTAATGCGCTGCGCGCGCGCGCGCTGCTCCACCTTCACATGAGCTC 1778
Db 1718 CTGTTGCAAGATTAATGCGCTGCGCGCGCGCGCGCTGCTCCACCTTCACATGAGCTC 1777
QY 1779 CTGTTGCAAGATTAATGCGCTTACACTCTTCTACCTGCGCGGCGCTTCATATTAACCT 1838
Db 1778 CTGTTGCAAGATTAATGCGCTTACACTCTTCTACCTGCGCGGCGCTTCATATTAACCT 1837
QY 1839 GAGCAGCTGTGAGACAGTGGCGCGCTGATTTTCAAAAGTCTCTGCGGAGGAGT 1898
Db 1838 GGACAAAGCTGTGAGACAGTGGCGCGCTGATTTTCAAAAGTCTCTGCGGAGGAGT 1897

QY 1899 TGAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1958
Db 1898 CTCGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1957
QY 1959 CACCATTCGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2018
Db 1958 CACCATTCGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2017
QY 2019 CTAGGCAATCTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2078
Db 2018 CTAGGCAATCTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2077
QY 2079 GTCTTAAAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2138
Db 2078 ATCTTGAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2137
QY 2139 CCGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2198
Db 2138 ACTGCGG-----GGACCTTTCCCGGCTGCGCCACCCAGAGGAGGAGGAGGAGGAGGAG 2184
QY 2199 CTGGGAGCAGTGAAGCAATGACCTTACAGATGCTGATGATGATGATGATGATGATGATGAT 2258
Db 2185 CTGGGAGCAGTGAAGCAATGACCTTACAGATGCTGATGATGATGATGATGATGATGATGAT 2237
QY 2259 CAGTGGCAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2297
Db 2238 CAGTGGCAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2276

RESULT 9
AX685731
LOCUS 2019 bp DNA linear PAT 29-MAR-2003
DEFINITION Sequence 3 from Patent WO02081691.
ACCESSION AX685731
VERSION AX685731.1 GI:29371740
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
Hobbs, H.H., Shan, B., Barnes, R. and Tian, H.
Abcg8 and abcg8: compositions and methods of use
Patent: WO 02081691-A 3 17-OCT-2002;
Tularik Inc. (US); BOARD OF REGENTS UNIVERSITY OF TEXAS SYSTEM
(US)

FEATURES
source
1..2019 Location/Qualifiers
1..2019 /organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
1..2019 /note="unnamed protein product; mouse ABCG8 (mABG8)"
/codon_start=1
/protein_id="CAD86571.1"
/db_xref="GI:29371741"
/translation="MAEKTEKTLQVNGVILQDASGLDLSFSSPSDNLSTPTSGOS
NLEVRDLTYVDIASOVPMFQLOAPRTKPMHSOSDSCEGLRNLSTKVSQOMLA
ITGGSGGASLADVTGKRGHGKMSQDITNGPSTPQIVLRKCAVAHRQDOLPN
LTVREGLAFIAQRPRPESQAKRQREDVIAELRLRCQATRVGNTVYRVSQGER
RVSIGVOLNMBGILIDEPSTGDSFTAHNLVTLRLAAGNLVLISLQPSDI
RLFLDLVLTMTGTPYLAGAOMQVYFTSIGHPCPSNPADFYDLTSLIDRSKER
EVAETKASQSLAFLKLVQGGFDDILAKAEKELMTSTHTVSTLTLOPTDCTAYELP
GMIBPSTLIRQINDRDLPTLILHSEKALMSLTIGFLYTGHAQQLSMDTAL
LPMIALIPENYILDVNSKCHERSMLYLEEDGLTAGPYFFAKITLDELPHCAVYI
ISYAMPITVMTNLRVPELFLAFILFLVWLVVCCRTMALASMLPTPFMSSFCALY
NSFYLTAGFMINDMLMIVPAISKLSLRFMCSSGIMQIOFNGHLYTTOIGNFTSIL
GDTMSADLNHPLPYAIYLIYIGTSYGFLLYIYLSLKLINOKSIDM"

Query Match

53.6%; Score 1430; DB 6; Length 2019;

	Query Match	27.13;	Score 724;	DB 9;	Length 127066;
	Best Local Similarity	90.08;	Pred. No. 1,1e-149;		
	Matches 824;	Conservative 0;	Mismatches 5;	Indels 87;	Gaps 1;
OY	1841 GCACGCTGTGACAGTGGCCGCGCTGGATTTCCAAAGTGCTCTTCGCGGAGGCTTTTG	1900			
Db	59498 GCTGTCGTGCCAGTGGCCCGCGTGGATTTCCAAAGTGCTCTTCGCGGAGGCTTTTG	5955			
OY	1901 AAGGCGTGATGAAATTTCACTTTCAGTCAGCGAAGAACTTATAAATGCTCTCGGGAACCTCA	1960			
Db	59558 AAGGCGTGATGAAATTTCACTTTCAGTCAGCGAAGAACTTATAAATGCTCTCGGGAACCTCA	5961			
OY	1961 CCATCGCGGTCCTCAGAGATATAA				
Db	59618 CCATCGCGGTCCTCAGAGATATAAGTAAAGCGGGAGAGCCCTCGGGTCTTAAATTAATTTGAC	59677			
OY	1984				
	-----ATCCTCAGTG	1993			
Db	59678 GTCCGCGCTTCATCTCTCTCATGAGCCCACTGATGTCTGTGTCACATCTCTCAGTG	59733			
OY	1994 CCATGAGAGTGAGACTCTACCTCTCTACGCGCATTTACCTCATCTGTCATTTGGCTCAGCG	2053			

Db	59738	CCATGGAGACTGGACTCGTACCCCTTCTTACGCCATCTTACCTCATGTCATATTGGCTCAGCG	59792
OY	2054	GTGGCTTCATGGCTCCTGTACTACTGTCTTCTTAAAGTTTCATCAAAAGAAACCAAGTCAAG	2113
Db	59798	GTGGCTTCATGGCTCCTGTACTACTGTCTTCTTAAAGTTTCATCAAAAGAAACCAAGTCAAG	59855
OY	2114	ACTGTTGATTTACGCCAGAGCTCTGCCCCGTGTGGGGGACCTTGACAGACCCCTTCACCT	2173
Db	59858	ACTGTTGATTTACGCCAGAGCTCTGCCCCGTGTGGGGGACCTTGACAGACCCCTTCACCT	59917
OY	2174	GCACCTCCCTCCTCAGAGAGCCCTTCCCTGGGGAGACAGTGAAGACATACCTTACACATGCT	2233
Db	59918	GCACCTCCCTCCTCAGAGAGCCCTTCCCTGGGGAGACAGTGAAGACATACCTTACACATGCT	59977
OY	2234	CAGCTACATCCGGGCCAGGGTGTCTGCTGAGTGGCAGACCCAGCCAGAGATGGGAGTGAAGA	2293
Db	59978	CAGCTACATCCGGGCCAGGGTGTCTGCTGAGTGGCAGACCCAGCCAGAGATGGGAGTGAAGA	60037
OY	2294	TAAAGACAGTTCGAAAGGGATTTTCTCTCACTGGCAGAGAGACTGCGATGACTGGAGAGAAA	2353
Db	60038	TAAAGACAGTTCGAAAGGGATTTTCTCTCACTGGCAGAGAGACTGCGATGACTGGAGAGAAA	60097
OY	2354	CCTGCACATCCGGTGGCAGCTTACAAAGCTTGAATTTATTTCTTTGATATGCAATTTAT	2413
Db	60098	CCTGCACATCCGGTGGCAGCTTACAAAGCTTGAATTTATTTCTTTGATATGCAATTTAT	60157
OY	2414	AGGCACTCGATATAGAGTGGAGGAGCAATACAGTACAGTACAGATTTGGCTTATCTTCA	2533
Db	60158	AGGCACTCGATATAGAGTGGAGGAGCAATACAGTACAGTACAGATTTGGCTTATCTTCA	60217
OY	2474	GAATTTGTTGGAACCTGGAGGGAACATTAACAGTACAGTACAGATTTGGCTTATCTTCA	2533
Db	60218	GAATTTGTTGGAACCTGGAGGGAACATTAACAGTACAGTACAGATTTGGCTTATCTTCA	60277
OY	2534	GGGGCCCCCAGACATCTCGTGTGAGCCACACATCAATACAGAAAGTACCTAAGATGACAG	2593
Db	60278	GGGGCCCCCAGACATCTCGTGTGAGCCACACATCAATACAGAAAGTACCTAAGATGACAG	60337
OY	2594	CAAGATGCGAATCCCTTTTGTGTGGGGGTCTATGGGCTCCAAAGGCCAAGCTGAACAT	2653
Db	60338	CAAGATGCGAATCCCTTTTGTGTGGGGGTCTATGGGCTCCAAAGGCCAAGCTGAACAT	60397
OY	2654	TAAAAATGTATTGAGC 2669	
Db	60398	TAAAAATGTATTGAGC 60413	
RESULT 11			
AC087053	182261 bp	DNA	linear HTG 27-MAR-2003
LOCUS	AC087053	Homo sapiens chromosome 2 clone RP11-959M3 map 2, 9 unordered	
DEFINITION			
ACCESSION	AC087053		
VERSION	AC087053.13	GI:25140148	
KEYWORDS	HTGS_PHASE1; HTGS_FULLTOP; HTGS_CANCELLED.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE			
AUTHORS	Bliren, B., Nisbaum, C. and Lander, E.		
TITLE	1 (bases 1 to 182261)		
JOURNAL	Homo sapiens chromosome 2, clone RP11-959M3		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 182261)		
	Bliren, B., Nisbaum, C., Lander, E., Allen, N., Anderson, S.,		
	Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A.,		
	Camarata, J., Campolano, A., Choebe, Y., Colangelo, M., Collins, S.,		
	Collymore, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J. S.,		
	Dodge, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,		
	Gardyna, S., Glade, S., Goyette, M., Graham, L., Grand-Pierre, N.,		
	Hagos, B., Hearford, A., Horton, L., Huine, W., Iliev, I., Johnson, R.,		
	Jones, C., Karatas, A., Lakoque, K., Lamazares, R., Landers, T.,		
	Lehoczky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P.,		

OY 2594 CAGATGCCATCCCTTTTGTGCGGGTCATGGGCTCAACCAAGCAAGTCAACAT 2653
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 7710 CAGATGCCATCCCTTTTGTGCGGGTCATGGGCTCAACCAAGCAAGTCAACAT 7769
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 2654 TAAAAATGATTGAGC 2669
 ||||||||||||||||||
 DB 7770 TAAAAATGATTGAGC 7785
 ||||||||||||||||||
 RESULT 12
 AC108476 139342 bp DNA linear PRI 16-APR-2002
 LOCUS AC108476
 DEFINITION Homo sapiens BAC clone RP11-1413K20 from 2, complete sequence.
 AC108476
 VERSION AC108476.5 GI:19807988
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo;
 REFERENCE 1 (bases 1 to 139342)
 AUTHORS Sulston, J.E. and Waterston, R.
 TITLE Toward a complete human genome sequence
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 PUBMED 9847074
 REFERENCE 2 (bases 1 to 139342)
 AUTHORS Harting, C., Haekenson, W. and Doeber, A.
 TITLE The sequence of Homo sapiens BAC clone RP11-1413K20
 JOURNAL Unpublished (2001)
 REFERENCE 3 (bases 1 to 139342)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (27-JAN-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 4 (bases 1 to 139342)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (20-FEB-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 5 (bases 1 to 139342)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (29-MAR-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 6 (bases 1 to 139342)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (16-APR-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Mar 29, 2002 this sequence version replaced gi:18767626.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@wustl.edu
 ----- Summary Statistics
 Center project name: H_NH1413K20

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.
 McPherson, Department of Genetics, Washington University, St. Louis
 MO. For additional information about the map position of this
 sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCT-11 human BAC library was made from the blood of one male
 donor, as described by Osoegawa, K., Woon, P. Y., Zhou, B., Firenzen, E.,
 Tatenno, M., Catanesse, J. J., and de Jong, P. J. (1998) An improved
 approach for construction of bacterial artificial chromosome
 libraries. Genomics 51:1-8. The clone may be obtained either from
 Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong
 and coworkers at <http://www.chori.org>
 VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-489K22, 2000 bp overlap.
 Actual end is at base position 139342 of RP11-1413K20.

The region between 132012 and 132017 is covered only by a PCR
 product of clone DNA.

FEATURES

source
 1. 139342
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2"
 /clone="RP11-1413K20"
 /clone_lib="RPCT-11"
 55. .653
 /note="match to EST AA203458 (NID:g1799169) zx58b04.r1"
 misc_feature
 93. .286
 /note="match to EST AV689089 (NID:g10290952)"
 misc_feature
 93. .286
 /note="match to Mus musculus EST A1597378 (NID:g4606426)
 vj29c06.y1"
 misc_feature
 93. .279
 /note="match to EST AV660973 (NID:g9881987)"
 misc_feature
 318. .653
 /note="match to EST R00405 (NID:g750141) ye71e05.r1"
 misc_feature
 372. .633
 /note="match to EST R00405 (NID:g750141) ye71e05.r1"
 /note="similar to Homo sapiens EST T97887 (NID:g747232)
 ye58h05.r1"
 misc_feature
 706. .708
 /note="match to EST R00405 (NID:g750141) ye71e05.r1"
 706. .707
 /note="similar to Homo sapiens EST T97887 (NID:g747232)
 ye58h05.r1"
 repeat_region
 847. .1139
 /rpt_family="Alu"
 misc_feature
 1867. .2047
 /note="match to EST T39945 (NID:g647612) ya13g04.r1"
 repeat_region
 2234. .2616
 /rpt_family="L2"
 misc_feature
 2983. .3121
 /note="match to EST AV689089 (NID:g10290952)"
 misc_feature
 2983. .3121
 /note="similar to Mus musculus EST A1597378 (NID:g4606426)
 vj29c06.y1"
 misc_feature
 3044. .3121
 /note="match to EST T86384 (NID:g714736) yd77b08.r1"
 misc_feature
 4099. .4304
 /note="match to EST T86384 (NID:g714736) yd77b08.r1"
 misc_feature
 4099. .4283
 /note="match to EST AV689089 (NID:g10290952)"
 4401. .4618
 /note="similar to Mus musculus EST BF162656


```

Query Match      27.18; Score 722.4; DB 9; Length 139342;
Best Local Similarity   89.88; Pred. No. 2.5e-149;
Matches 823; Conservatly    0; Mismatches    6; Indels    87; Gaps    1;

1841 GCACCGCTGTGAGCAGTCCGGCGGATTCGCCAAGAAGTGCTCTCTCTGCGGTGCTTTTGC
||||| |
misc_feature      (NID:g11042879) "
4405..4454
/note="match to EST T86384 (NID:g714736) yd77Db08.r1"
4724..5110
/note="similar to Homo sapiens EST AV656623
(NID:g9877637)"
5075..5204
/note="similar to Mus musculus EST BF162656
(NID:g11042879)"
5495..5657
/rpt_family="MIR"
5673..5767
/rpt_family="MIR"
5774..5813
/rpt_family="MIR"
5816..5958
/rpt_family="Alu"
5976..6091
/rpt_family="MIR"
6162..6485
/rpt_family="Alu"
6351..6373
/note="match to EST AA28345 (NID:g1849916) nc39d04.si"
6352..6364
/note="match to EST AI431309 (NID:g4302284) ar55B01.x1"
6352..6364
/note="match to EST AI469772 (NID:g4331862) tm20F11.x1"
6353..6367
/note="match to EST AI241685 (NID:g3837082) qu70f06.x1"
6568..6707
/note="similar to Mus musculus EST BF162656
(NID:g11042879)"
6649..6707
/note="similar to Mus musculus EST BB598373
(NID:g16450340)"
7229..7528
/rpt_family="Alu"
7940..8549
/note="similar to EST BM725726 (NID:g19047059)"
8169..8305
/note="similar to Mus musculus EST BF162656
(NID:g11042879)"
8169..8301
/note="similar to Mus musculus EST BB598373
(NID:g16450340)"
8500..8529
/rpt_family="AT_rich"
8540..8868
/rpt_family="Alu"
8870..9180
/rpt_family="Alu"
10493..10636
/rpt_family="MIR"
11195..11376
/rpt_family="MERL_type"
11377..11658
/rpt_family="Alu"
11659..11799
/rpt_family="MERL_type"
11955..12053
/note="similar to Mus musculus EST BB598373
(NID:g16450340)"
11994..12053
/note="similar to Mus musculus EST AA239884 (NID:g18639223)
mx81d01.r1"
12086..12109
repeat_region

```

	Db	57732	GCTGTCTGTCTCCAGTGTCCCGCGGTGATTTCCAAAGTGTCTTCCGTGCGGTGTGTTTG	57791
QY	1901	AAGGGCTGATGAAAGATTCTAGTTCAGCAGAGAAGAACTATATAATGCTCTCGGGAACTCA	1960	
Db	57792	AAGGGCTGATGAAAGATTCTAGTTCAGCAGAGAAGAACTATATAATGCTCTCGGGAACTCA	57851	
QY	1961	CCATCGGGGCTCTCGAGGATTA-----	1983	
Db	57852	CCATCGGGGCTCTCGAGGATTAAGATAAGCGGGGAAGGCTCGGGTTCTATAATATATGAC	57911	
QY	1984	-----ATCCCTCAGTG	1993	
Db	57912	GTCCGGCTTTCATCCCTCCTCATAGAGCCCACTCATGTCTGTGTCTCCAGATCTCTAGTG	57971	
QY	1994	CCATGAGAGCTGAGACTGCTACCCCTCTCTACGCCATCTACCTCATCTGATTTGGCTTCAGTG	2053	
Db	57972	TCATGAGAGCTGAGACTGCTACCCCTCTCTACGCCATCTACCTCATCTGATTTGGCTTCAGTG	58031	
QY	2054	GTGGCTTCATGATGTCTGTCTACTACGTGTCTTAAAGTTTCATCAAGAAAGAAAGCAAGTCAAG	2113	
Db	58032	GTGGCTTCATGATGTCTGTCTACTACGTGTCTTAAAGTTTCATCAAGAAAGAAAGCAAGTCAAG	58091	
QY	2114	ACTGTGATTTACAGCCAGACGCTGTGGCCGCTGTGGGGAGACCTGAGACAGACCTTCAACT	2173	
Db	58092	ACTGTGATTTACAGCCAGACGCTGTGGCCGCTGTGGGGAGACCTGAGACAGACCTTCAACT	58151	
QY	2174	GCACTCCCTCTCTGAGAGGACCCCTTCTCTGGGGACAGTGAAGACAAATGACCTTCAAGATGCT	2233	
Db	58152	GCACTCCCTCTCTGAGAGGACCCCTTCTCTGGGGACAGTGAAGACAAATGACCTTCAAGATGCT	58211	
QY	2234	CAGCTACATCCGGCCCAAGGGTGTCTGACATGAGCAGACAGCCAGCAGAGTGGCAATAGAA	2293	
Db	58212	CAGCTACATCCGGCCCAAGGGTGTCTGACATGAGCAGACAGCCAGCAGAGTGGCAATAGAA	58271	
QY	2294	TAAAGACAGTCAAAAGGAAATTTCTGCTACTGGCAGAGACATGCGATGACTGTGGGAGAAA	2353	
Db	58272	TAAAGACAGTCAAAAGGAAATTTCTGCTACTGGCAGAGACATGCGATGACTGTGGGAGAAA	58331	
QY	2354	CTGCACTCGTGGTGGACACTACAAAGTTGTCTAAATTTATTTTCTTTGATATGATTTATAT	2413	
Db	58332	CTGCACTCGTGGTGGACACTACAAAGTTGTCTAAATTTATTTTCTTTGATATGATTTATAT	58391	
QY	2414	AGGCAACTGATATAGAGATGGAGCAACTAGAGATGAATTTGGTAGCTAGACTGTGCAG	2473	
Db	58392	AGGCAACTGATATAGAGATGGAGCAACTAGAGATGAATTTGGTAGCTAGACTGTGCAG	58451	
QY	2474	GAATTTGTGGAACTCTGGAGGAAACAATACACTACTACTAGCAATTTGGCTTCAATCTTCCA	2533	
Db	58452	GAATTTGTGGAACTCTGGAGGAAACAATACACTACTACTAGCAATTTGGCTTCAATCTTCCA	58511	
QY	2534	GGGGCCCCACACTCCGTGTGTGAGCCACATCAATACAGAAAGTGAAGTGAAGTGAACAG	2593	
Db	58512	GGGGCCCCACACTCCGTGTGTGAGCCACATCAATACAGAAAGTGAAGTGAAGTGAACAG	58571	
QY	2594	CAGATGCAATCCCTCTTTTGTGTGTGGGCTCATGGGCTCCAAAAGCCAAAGTGAACAT	2653	
Db	58572	CAGATGCAATCCCTCTTTTGTGTGTGGGCTCATGGGCTCCAAAAGCCAAAGTGAACAT	58631	
QY	2654	TAAAAATGTAATGAGC	2669	
Db	58632	TAAAAATGTAATGAGC	58647	
RESULT 13				
F351812S13				
LOCUS				
F351812S13 2201 bp DNA linear PRI 10-AUG-2001				
DEFINITION Homo sapiens sterol-in-2 (ABCG8) gene, exon 13 and complete cds.				
ACCESSION AF351824				
VERSION AF351824.1 GI:15146443				
KEYWORDS				
SEGMENT				
SOURCE 13 of 13				
ORGANISM Homo sapiens (human)				
Homosapiens				

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Lu, K., Lee, M.H., Hazard, S., Brooks-Wilson, A., Hidaka, H., Kojima, H., Ose, L., Stalenhoef, A.F., Mettenen, T., Bjorkhem, I., Bruckert, E., Pandya, A., Brewer, H.B. Jr., Salen, G., Dean, M., Srivastava, A. and Patel, S.B.

TITLE

Two genes that map to the STSL locus cause sitosterolemia: genomic structure and spectrum of mutations involving sterol-1 and sterol-2, encoded by ABCG5 and ABCG8, respectively

JOURNAL
MEDLINE
PUBMED

Am. J. Hum. Genet. 69 (2), 278-290 (2001)

REFERENCE
AUTHORS

Lu, K.
Submitted (21-FEB-2001) Division of Endocrinology, Diabetes and Medical Genetics, Medical University of South Carolina, 114 Doughty St, STB 541, Charleston, SC 29403, USA

FEATURES

Location/Qualifiers

1..2201
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/map="between D2S2294 and D2S2298"
/clone="108162; 32814"

gene

/cell_type="ES cell"
order(AFG51812.1:<1227..2809,AFG51813.1:1..4665,
AFG51814.1:1..2368,AFG51815.1:1..1323,AFG51816.1:1..300,
AFG51817.1:1..660,AFG51818.1:1..203,AFG51819.1:1..888,
AFG51820.1:1..685,AFG51821.1:1..884,AFG51822.1:1..1292,
AFG51823.1:1..859,1..>182)

mRNA

/gene="ABCG8"
join(AFG51812.1:<1227..1289,AFG51813.1:3941..4042,
AFG51814.1:924..1080,AFG51815.1:674..912,
AFG51816.1:121..253,AFG51817.1:66..335,AFG51818.1:1..163,
AFG51819.1:48..128,AFG51820.1:310..509,
AFG51821.1:243..319,AFG51822.1:101..368,
AFG51823.1:689..816,45..>182)

CDS

/product="sterolin-2"
join(AFG51812.1:1227..1289,AFG51813.1:3941..4042,
AFG51814.1:924..1080,AFG51815.1:674..912,
AFG51816.1:121..253,AFG51817.1:66..335,AFG51818.1:1..163,
AFG51819.1:48..128,AFG51820.1:310..509,
AFG51821.1:243..319,AFG51822.1:101..368,
AFG51823.1:689..816,45..182)

/note="ABCG8"
/gene="ABCG8"
/codon_start=1
/product="sterolin-2"
/protein_id="AAK84663.1"

/db_xref="GI:1514644"

/translation="MAGKAAEEELPKGATPDNSGLDRIFFSESDNSLYPTSGOP
NTLEAYDNCQVDASQVPEFQIAQRPMTSPSCQSCLEGIONLSFKVRSQMLA
ITGSGGRASLLVITRGHGKIRKSOQINQSPSSPOLVRCVCAVNRHNDLPN
LTVETLAFIAQMLPRTFSOARDKREVDYIAELRTKADTVGVNAYVGLSGGER
FRVIGVGLNNPGLIIDEPTSGLDSTANLAKTSLAKNRLVYISIHORSDI
FRFDVLVMTSGPIYLGAAOHVYETAIGCCPRSNADYVDLSTDRSRDO
ELATREKQASIALEFLEKVRIDDELFKAEKRDDEPTCVESYPLDINCPSTKAP
GAVOQFTLIRALISNDRDPTLLIHGACACMLTGTGFIYFGHSIQLSFMDTAL
LPMIGALIPNVILDVISKCYSERAMLYELEDELVYTFPAKITIGEEHAYII
ITGMPTWMLANRGLOPFLHPLVLMVCCIMLAAALPTFPMASFSFNALY
NSFIAGGFMINLSMTVPMAMISKVSFLKWCPEGLMKIGSRRTYKMGGLTIAVS
GDKILSMEIDSYPLVAYILVIGLSSGFVLYVSLRFIKRQSDM"

45..>182
/gene="ABCG8"
/number=13

exon

BASE COUNT
ORIGIN

543 a 536 c 538 g 580 t 4 others

Query Match

24.9%; Score 663.6; DB 9; Length 2201;

Best Local Similarity 99.3%; Pred. No. 3,4e-136;
Matches 685; Conservative 2; Mismatches 1; Indels 2; Gaps 2;

QY	1982	AAATCTGAGTGCATGAGGAGCTGAGTGTACCTCTCTACAGCCATCTACCCATGCTCA	2041
DB	43	AGATCTCTGAGTGCATGAGGAGCTGAGTGTACCTCTCTACAGCCATCTACCCATGCTCA	102
QY	2042	TTGGCTCAGCGGTGGCTTCATGCTCTGCTTACTAGCTGCTTAAGGTTCATCAACAGA	2101
DB	103	TTGGCTCAGCGGTGGCTTCATGCTCTGCTTACTAGCTGCTTAAGGTTCATCAACAGA	162
QY	2102	AACCAAGTAAAGCTGCTGATTCACGCCAGACGCTGCTGCCGTGGTGGGACCTGAGCA	2161
DB	163	AACCAAGTAAAGCTGCTGATTCACGCCAGACGCTGCTGCCGTGGTGGGACCTGAGCA	222
QY	2162	GACCTTCAGTGCATCTCTCTCTGAGGAGCCCTTCCTGGGAGACATGACATATAC	2221
DB	223	GACCTTCAGTGCATCTCTCTCTGAGGAGCCCTTCCTGGGAGACATGACATATAC	282
QY	2222	CCTACAGATGCTCAGCTACATCCGCGCCAGGTGCTGACAGTGCACAGCCAGCAG	2281
DB	283	CCTACAGATGCTCAGCTACATCCGCGCCAGGTGCTGACAGTGCACAGCCAGCAG	342
QY	2282	ATGGCAGTGAATTAAGACAGTGCAGAAAGGATTTCTCTACTGCGAGAGACTGCATG	2341
DB	343	ATGGCAGTGAATTAAGACAGTGCAGAAAGGATTTCTCTACTGCGAGAGACTGCATG	402
QY	2342	ACTGGGAGAAACCTGCACCTCGCTGGGACCTCAACGCTTCTGTAATTTCTTTGAT	2401
DB	403	ACTGGGAGAAACCTGCACCTCGCTGGGACCTCAACGCTTCTGTAATTTCTTTGAT	462
QY	2402	ATGCATTTATATAGCACTCGATATAGATGGAGCAACATAGGAATTAATTTGGTGC	2461
DB	463	ATGCATTTATATAGCACTCGATATAGATGGAGCAACATAGGAATTAATTTGGTGC	522
QY	2462	TAGACTGTCAGGAATTTGTGGACCTGGAGGAAATTAACATAGTACAGATTTGG	2521
DB	523	TAGACTGTCAGGAATTTGTGGACCTGGAGGAAATTAACATAGTACAGATTTGG	582
QY	2522	CTTCATCTCCAGGGGCCCCACATCTCGTGTGAGCCACATCAATACAGAAATGACCT	2581
DB	583	CTTCATCTCCAGGGGCCCCACATCTCGTGTGAGCCACATCAATACAGAAATGACCT	642
QY	2582	AAGATGTACAGCAAGATGCCATCTCTTTTGTGTGGGTCATGGGCTCCAAAAGC	2640
DB	643	NAAAGTGTACAGCAAGATGCCATCTCTTTTGTGTGGGTCATGGGCTCCAAAAGC	702
QY	2641	CAACGTGAACAA-TTAAAAATGTATTTGAGC 2669	
DB	703	CAACGTGAACAA-TTAAAAATGTATTTGAGC 732	
RESULT 14	F351812S06		
LOCUS	F351812S06	660 bp.	DNA linear PRI 10-AUG-2001
DEFINITION	Homo sapiens sterol-2 (ABCG8) gene, exon 6.		
ACCESSION	AF351817		
VERSION	AF351817.1	GI:15146436	
KEYWORDS	6 of 13		
SEGMENT			
SOURCE	Homo sapiens (human)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 660)		
AUTHORS	Lu, K., Lee, M.H., Hazard, S., Brooks-Wilson, A., Hidaka, H., Kojima, H., Ose, L., Stalenhoef, A.F., Mettenen, T., Bjorkhem, I., Bruckert, E., Pandya, A., Brewer, H.B. Jr., Salen, G., Dean, M., Srivastava, A. and Patel, S.B.		
TITLE	Two genes that map to the STSL locus cause sitosterolemia: genomic structure and spectrum of mutations involving sterol-1 and sterol-2, encoded by ABCG5 and ABCG8, respectively		
JOURNAL	Am. J. Hum. Genet. 69 (2), 278-290 (2001)		

MEDLINE 21344600
 PUBMED 11452359
 REFERENCE 2 (bases 1 to 660)
 AUTHORS Lu, K.
 TITLE Direct Submission
 JOURNAL Submitted (21-FEB-2001) Division of Endocrinology, Diabetes and Medical Genetics, Medical University of South Carolina, 114 Doughty St., STB 541, Charleston, SC 29403, USA

FEATURES
 source location/Qualifiers
 1. 660
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="between D2S2294 and D2S2298"
 /clone="1081G2, 32814"
 /cell_type="ES cell"
 66. 335
 /gene="ABCG8"
 /number=6

exon
 BASE COUNT 158 a 184 c 172 g 146 t
 ORIGIN

Query Match 10.3%; Score 274.2; DB 9; Length 660;
 Best Local Similarity 97.2%; Pred. No. 6.8e-50;
 Matches 279; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 781 CTGTGGAACCCAGGATTCCTATTCTGACGAAACCCACCTCTGGCTCCACAGCTTCACA 840
 |||||
 DB 53 CTGTGTGGAAGAAATCCCTATTCTGACGAAACCCACCTCTGGCTCCACAGCTTCACA 112
 |||||

OY 841 GCCCAACCTGGTGAAGACCTGTCCAGGCTGGCCAAAGGCAACGGGCTGCTCAGC 900
 |||||
 DB 113 GCCCAACCTGGTGAAGACCTGTCCAGGCTGGCCAAAGGCAACGGGCTGCTCAGC 172
 |||||

OY 901 TCCCTCCACAGGCTGCTGTGACATCTTCAGGCTGTTGATCTCTCTCTGATGACG 960
 |||||
 DB 173 TCCCTCCACAGGCTGCTGTGACATCTTCAGGCTGTTGATCTCTCTCTGATGACG 232
 |||||

OY 961 TCTGGACACCCATCTACCTTAGAGGGGGCCACACATGATGCTCAGATTTTCACAGCCATC 1020
 |||||
 DB 233 TCTGGACACCCATCTACCTTAGAGGGGGCCACACATGATGCTCAGATTTTCACAGCCATC 292
 |||||

OY 1021 GGCTACCCCTGCTCGTACAGCAATCCGCTGACTTCTATGTGA 1067
 |||||
 DB 293 GGCTACCCCTGCTCGTACAGCAATCCGCTGACTTCTATGTGA 339
 |||||

RESULT 15
 F351812S11 1292 bp DNA linear PRI 10-AUG-2001
 LOCUS Homo sapiens steroln-2 (ABCG8) gene, exon 11.
 DEFINITION AF351822
 ACCESSION AF351822.1 GI:15146441
 VERSION
 KEYWORDS
 SEGMENT
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 11 of 13
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS
 1 (bases 1 to 1292)
 Lu, K., Lee, M.H., Hazard, S., Brooks-Wilson, A., Hidaka, H., Kojima, H.,
 Ose, L., Stalenhoef, A.F., Miettinen, T., Bjorkhem, I., Bruckert, E.,
 Pandya, A., Brewer, H.B. Jr., Salen, G., Dean, M., Sirtavastava, A. and
 Patel, S.B.
 TITLE Two genes that map to the STSL locus cause sitosterolemia: genomic
 structure and spectrum of mutations involving steroln-1 and
 steroln-2, encoded by ABCG5 and ABCG8, respectively
 JOURNAL Am. J. Hum. Genet. 69 (2), 278-290 (2001)
 MEDLINE 21344600
 PUBMED 11452359
 REFERENCE 2 (bases 1 to 1292)
 AUTHORS Lu, K.

TITLE Direct Submission
 JOURNAL Submitted (21-FEB-2001) Division of Endocrinology, Diabetes and Medical Genetics, Medical University of South Carolina, 114 Doughty St., STB 541, Charleston, SC 29403, USA

FEATURES
 source location/Qualifiers
 1. 1292
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="between D2S2294 and D2S2298"
 /clone="1081G2, 32814"
 /cell_type="ES cell"
 101. 368
 /gene="ABCG8"
 /number=11

exon
 BASE COUNT 233 a 405 c 317 g 327 t 10 others
 ORIGIN

Query Match 10.2%; Score 273.4; DB 9; Length 1292;
 Best Local Similarity 86.7%; Pred. No. 9.9e-50;
 Matches 301; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

OY 1586 AGATCCCGGGGAGCTTCGGAGCAGCTGCTCAATCATCATACGATGAGATGCCACCT 1645
 |||||
 DB 99 AGATCCCGGGGAGCTTCGGAGCAGCTGCTCAATCATCATACGATGAGATGCCACCT 158
 |||||

OY 1646 ACTGGCTGGCCAACTGAGGCCAGGCTCCAGCCCTCCGCTGCTGCTGCTGCTGCTG 1705
 |||||
 DB 159 ACTGGCTGGCCAACTGAGGCCAGGCTCCAGCCCTCCGCTGCTGCTGCTGCTGCTG 218
 |||||

OY 1706 GCGTGTGCTCTCTGCTGAGATTAAGGCTGAGCCGCGCGGCGCTGCTGCCACCT 1765
 |||||
 DB 219 GCGTGTGCTCTCTGCTGAGATTAAGGCTGAGCCGCGCGGCGCTGCTGCCACCT 278
 |||||

OY 1766 TCCACATGGGCTCTCTTCCAGCAATGCCCTCAACATCTCTACCTGCGGGGGCT 1825
 |||||
 DB 279 TCCACATGGGCTCTCTTCCAGCAATGCCCTCAACATCTCTACCTGCGGGGGCT 338
 |||||

OY 1826 TCATGATTAACCTGAGCAGCTGTGACAGTGGCCCGCTGATTTCCAAAGTGTCTCC 1885
 |||||
 DB 339 TCATGATTAACCTGAGCAGCTGTGACAGTGGCCCGCTGATTTCCAAAGTGTCTCC 398
 |||||

OY 1886 TCGGCTGTGCTTTTGAAGGCTGATGAAGATTCAATTCAGACAGAGA 1932
 |||||
 DB 399 TCGGCTGTGCTTTTGAAGGCTGATGAAGATTCAATTCAGACAGAGA 445
 |||||

Search completed: July 27, 2003, 05:52:39
 Job time : 9494.57 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2003, 23:50:29 ; Search time 4963.41 Seconds
(without alignments)
13069.363 Million cell updates/sec

Title: US-09-989-981a-7

Sequence: 1 gtgtccctgtccaggaac.....caatlaaatgtatgagc 2669

Scoring table: IDENTITY_NTC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthm:*
3: em_estln:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rnd:*
26: em_gss_phg:*
27: em_gss_vrt:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1511.6	56.6	3623	11	AK004871	AK004871 Mus muscu
2	1286.2	48.2	2417	11	AK050938	AK050938 Mus muscu
3	681.8	25.5	691	2	BSM072006	BSM072006 Homo sapi
4	460.6	17.3	849	12	BI330745	BI330745 602982409

5	370.4	13.9	549	10	BE660076	BE660076 maa27c08.
6	361.4	13.5	583	14	BY705076	BY705076 BY705076
7	355.8	13.3	457	14	T91380	T91380 yd53b02.s1
8	332.4	12.5	334	2	BSM072530	BSM072530 Homo sapi
9	331.2	12.4	510	10	BSM10072	BSM10072 BSM10072
10	323	12.1	511	9	AI157365	AI157365 u45h01.y
11	318.4	11.9	564	14	T84531	T84531 u45b02.r1
12	309.8	11.6	500	9	AI151811	AI151811 u46c10.y
13	276.8	10.4	463	9	AA537862	AA537862 v35a03.r
14	276	10.3	781	14	CB502603	CB502603 ssalimges50
15	234.6	8.8	398	9	AI597406	AI597406 v35a03.y
16	226.6	8.5	586	11	AK008188	AK008188 Mus muscu
17	226.4	8.5	581	14	BY708144	BY708144 BY708144
18	202.8	7.6	435	13	BX099922	BX099922 BX099922
19	198.4	7.4	435	9	AI574075	AI574075 u167h11.y
20	196.4	7.4	821	29	BZ650554	BZ650554 OGCBA897C
21	149.2	5.6	523	6	AU195806	AU195806 Porphyra
22	145.4	5.4	447	12	BI145065	BI145065 602909138
23	144.2	5.4	936	10	BF162656	BF162656 601769307
24	141.6	5.3	823	14	CB649273	CB649273 OSJNEB13B
25	141	5.3	833	10	BF620684	BF620684 HVSMEC002
26	138.8	5.2	566	6	AU192726	AU192726 Porphyra
27	138	5.2	560	6	AU192604	AU192604 Porphyra
28	137.8	5.2	754	14	CB627408	CB627408 OS1EB02F
29	137	5.1	791	13	BO743924	BO743924 WHEA109.G
30	135.8	5.1	854	28	BH703879	BH703879 BOHUR247F
31	135	5.1	793	14	CB620418	CB620418 OS1EA05J
32	134.4	5.0	397	14	CA024091	CA024091 HZ48F10F
33	132	4.9	581	12	BZ245824	BZ245824 BZ245824
34	132	4.9	798	29	BZ531631	BZ531631 OGAFH497C
35	132	4.9	803	29	BZ986050	BZ986050 PUGKE39FD
36	132	4.9	822	29	BZ731642	BZ731642 OGCRCY267C
37	132	4.9	746	14	CB627374	CB627374 OGCRCY37M
38	131.4	4.9	746	14	CB627374	CB627374 OS1EB02E
39	126.6	4.7	816	10	BG299523	BG299523 HVSMEC002
40	126.4	4.7	274	14	R89160	R89160 yq02a04.r1
41	125.8	4.7	490	29	BZ614705	BZ614705 1946b01.g
42	124.4	4.7	867	14	CD382895	CD382895 PTM08065
43	123.6	4.6	841	14	CB683273	CB683273 OSJNEF11H
44	123.2	4.6	563	9	AV835407	AV835407 AV835407
45	122.6	4.6	812	29	BZ537267	BZ537267 OGA62867C

ALIGNMENTS

RESULT 1
LOCUS AK004871
DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:1300003C16 product:ATP-BINDING CASSETTE, SUB-FAMILY G, MEMBER 8 (STEROLIN-2) homolog [mus musculus], full insert sequence.
ACCESSION AK004871
VERSION AK004871.1 GI:12836380
KEYWORDS HTG; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
PUBMED 10349636
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374

REFERENCE	POBMED	11042159
TITLE	SHibata,K., Itoh,M., Alzawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,Y., Nishi,K., Katsuna,T., Tashiro,H., Itoh,M., Suni,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Saitaguchi,S., Ikegami,T., Kashiwagi,K., Fujimake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,K., Yoneda,Y., Ishikawa,T., Ozawa,Y., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.	
JOURNAL	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer	
MEDLINE	Genome Res. 10 (11), 1757-1771 (2000)	
POBMED	20530913	
REFERENCE	11076681	
AUTHORS	4	
TITLE	Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Aikawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamana,I., Saito,T., Okazaki,Y., Gotojori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Asbunier,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuno,Y., Nakaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staudt,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Momberters,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyoka,K., Wang,K.H., Wetz,C., Whitaker,C., Wilmung,L., Wyshaw-Bois,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohlski,S. and Hayashizaki,Y.	
JOURNAL	Functional annotation of a full-length mouse cDNA collection	
MEDLINE	Nature 409 (6821), 685-690 (2001)	
POBMED	21085660	
REFERENCE	11217851	
AUTHORS	5	
TITLE	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.	
JOURNAL	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	
MEDLINE	Nature 420, 563-573 (2002)	
POBMED	6 (bases 1 to 3623)	
REFERENCE	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arikawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Konda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Teijima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.	
JOURNAL	Direct Submission	
MEDLINE	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-rese@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)	
POBMED		
REFERENCE		
AUTHORS		
TITLE	Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.	
JOURNAL	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGGCGCCGCACTCGAGATTTTTTTTTTTTTTTTTTNN 3']. cDNA was	

prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adaptor of sequence 5' GAGGAGACAGACATCCAGAGTCATTTATTTATTAACCCCCCCC 3'. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

FEATURES	Location/Qualifiers
SOURCE	1. 3623
	/organism="Mus musculus"
	/mol_type="mrna"
	/strain="C57BL/6J"
	/db_xref="PANTOM.DB:1300003C16"
	/db_xref="MGI:1896857"
	/db_xref="taxon:10090"
	/clone="1300003C16"
	/sex="male"
	/tissue_type="liver"
	/clone.lib="RIKEN full-length enriched mouse cDNA library
	/dev_stage="adult"
CDS	69..2090
	/note="unnamed protein product: ATP-BINDING CASSETTE, SUB-FAMILY G, MEMBER 8 (STEROLIN-2) homolog [mus musculus (SMISSPROT Q9DBM0, evidence: FASTA, 9281D, 96.7%length, match=1196)]
	putative"
	/codon_start=1
	/protein_id="BAB23630.1"
	/db_xref="GI:12836381"
	/db_xref="MGI:1914720"
	/translation="MAEKTEKTEQLMNGTVIADASQIGLDSIFSESDNSLYTYSG SNLIEKRLIVQDVIASQVPMVEQLAQRIPWRSHSDSCSLGIRNLSFKRSQOM AIISSGCGRASLLDVIYIGRHGCKMKGQIINQPSPTPOLVRKCVAHROHDL NLVRETALFIAQMLPRTFSAQADKREVEYVIELRLQCAVTRNGVYVYVSGG RRRSYIGVLLMNPGLILDEPTEGSLDEFTAHNVLTSLAKGNRLVLI SLHPRS IRFEDVILMTSGPIYILGAQOOWOYETSIHGHCPRYSNPADPYDLTSDRSKI REVAVERAOSILAALFLEKVGQFDPELKKAKELNLTHTYSLTLTQDPTGAVEI PGMIKEPSTLIRROISNDPRDLPTILHGSERCLMSLIIGLYTGHAKOLSFMTAL LLPMGLIIPNVILDVNSRCHSEKSRMLYTELEDELVTAGPEYFAKILGELEPCAY IIVAPYILVNLNRPVELFLHLFLVWLVECCRTMLAASAMLPTRHMSSEFCNAL YNSVYILAGFMINDMNIWIPAMVISKISFLMCFSEGLMQIOFNGLIYTTQINFTES LGDPMISAMDLNLSHPLYAIYILVIGISYGFELYLSLTIKOKSIQDW"
polyA_signal	3605..3610
	/note="putative"
polyA_site	3623
	/note="putative"
BASE COUNT	847 a 1012 c 920 g 844 t
ORIGIN	
Query Match	56.6%; Score 1511.6; DB 11; Length 3623;
Best Local Similarity	77.0%; Pred. No. 0;
Matches 1965; Conservative	0; Mismatches 534; Indels 53; Gaps 8
QY	99 CATGCCGCGGAAGCGGCGACGAGAGAGAGAGGCGCTGCCGAAGGGCCACTCCCCAGAGATAC 158
DB	68 CATGCGCTAGAAAAACCAAGGAAGAACCCACTGTGAGTGGAGCTGACTTACAGATATC 127
QY	159 CTC---GGGCGCTCCAGATAGATTGTTCTCTCTGAAGTAGACACAGCCTTACTTATC 215
DB	128 TTCGAGGGGCTCCAGGACGACACTGTTCTCTCGGAAAGTAGACAAACATCTGTACTTAC 187
QY	216 CTAAGATGCGCAGCCCAACACCTCTGGAGGTGAGAGACTCAACTAATCAACAGGTGAGACTGCG 275
DB	188 CTAAGATGCTAGTCCCAACACTCTGGAGGTGAGAGATCTCACCTACCGAGGTGACATGCG 247
QY	276 CTCACAGTCCCTTGTTGAGAGAGCTGGCTCAGTTCAAAGATGCCCTGGACATCTCCAG 335
DB	248 CTCACAGTCCCTTGTTGAGAGAGCTGGCTCAGTTCAAAGATGCCCTGGACATCTCCAG 307
QY	336 CTCGCAATTTCTTGTGAGCTGGGCAATCAGAACCTAAAGTTCAAAGTGAAGTGGACA 395
DB	308 CAGCAAGATCCCTGTGAGCTGGGCAATCAGAACCTAAAGTGAAGTGGAGAGTGGACA 367
QY	396 GATCCTGACCATCATTAAGAGAGCTAGGTTGTGGGAGGACCTCTTGCTAGATGTGATAC 455

Db	368	GATGTGGCCATCATAGGAGGCTCAGGCTCGGGAGAGCCCTCACTACCTGACGTGATCAC	427
OY	456	TGGCCGAGGTACAGCGCGGCAAGATCAAAGTCAGGCCAGATCTGGATCAATGGGAGCCAG	515
Db	428	AGGCAGAGGCCACGCGTGGCAAGATGAATCAGACAAATTTGGATTAATGGGACACCAG	487
OY	516	CTGCGCTCAGCTGGTAGGAAGATGTGTGGCCACGTGCGCCAGCACACACAGCTGCTCCC	575
Db	488	TACGCTTCACCTGGTAGGAAGTGGCTGTGGCAATGTGGCGCACATGACCAACTGCTGCC	547
OY	576	CAACTTGACTGTGAGAGAGACTTGGGCGTTCAATGGCCAGATGGGCTGCCAGAACCTT	635
Db	548	CAACTGACCGTCAAGAGAGACCTTGCTTTCAATGGCCAGATGGGCTGCCAGAACCTT	607
OY	636	CTCCAGAGGCCAGCTGACAAAAGGGTGAAGAGAGTATGGCGAGCTGGGCGTTTAGCA	695
Db	608	CTCCAGAGGCCAGCTGACAAAAGGGTGAAGAGAGTATGGCGAGCTGGGCGTTTAGCA	667
OY	696	GTCGCGTGAACCCCGCTGGGCAACTATACGTGCGGGGGTGTGCGGGGGTAGCGCAG	755
Db	668	GTCGCGCAACAGCAGAGTGGGCAACAGTATGTACGTGGGGTGTCCGGGGTAGCGCGG	727
OY	756	GAGAGTCAGATTTGGGGTGGAGCTCGCTGTTGGGAACCCAGAAATCTTATTTGACGAC	815
Db	728	ACGAGTGAAGATTTGGGGTGGAGCTCGCTGTTGGGAACCCAGAAATCTTATTTGAC	787
OY	816	CACCTCTGGGCTGACAGCTTACAGCCCAACAACCTGTGTGAAGACCTTGTCCAGGGTGG	875
Db	788	CACCTCTGGGCTGACAGCTTACAGCCCAACAACCTGTGTGAAGACCTTGTCCAGGGTGG	847
OY	876	CAAAAGCAACCGGCTGTGCTATCTCCCTCCACAGCCCTGCTGACATCTTCAGGCT	935
Db	848	CAAGGGCAACAGGCTGTGCTATCTCCCTCCACAGCCCTGCTGACATCTTCAGGCT	907
OY	936	GTTTGAATCGGCTCTGATGACCGTGTGGCAACCCCATCTACTTAGGGGCGGCCAGCA	995
Db	908	ATTTGACCTGTGCTCTGATGACCGTGTGGCAACCCCATCTACTTAGGGGCGGCCAGCA	967
OY	996	CATGGTCCAGTATTTTACAGCCAGCGGCTACCCCTGTCTCGTACAGCAATCTGTGCTCA	1055
Db	968	AATGGTCAGTACTTCATTCATTTGGCCACCCCTGTCTCGTATAGCAACCTGTGCGGA	1027
OY	1056	CTTCTATGTGGACTGTACAGCAGATTGACAGGGCGACAGAGACAGAAATTTGGCCACAG	1115
Db	1028	CTTCTATGTGGACTGTACAGCAGATTGACAGGGCGACAGAGACAGAAATTTGGCCACAG	1087
OY	1116	GGAGAAGGCTGACTCTGCGCAGCCCTGTTTACAAAAGTGTGCTGACTTAGTACTT	1177
Db	1088	GGAGAAGGCTGACTCTGCGCAGCCCTGTTTACAAAAGTGTGCTGACTTAGTACTT	1147
OY	1176	TCTATGAAAGACAGACAGATCTTGACAGGAGCAACCTGTGTGAAGACGAGGTGAC	1233
Db	1148	TCTGTGAAAGCTGTGAGGCAAGAACTCAACAAGACACACACAGTCAGCTGACCTT	1207
OY	1236	CCCACTAGACACAACTGCTCTCCAGATCTTACGAAGATGCTGTGGGGGTGTACAGACTT	1295
Db	1208	CACACAGACACTGTACTGCTTGTGAGGCTGCCGGGATGATAGACACTT	1266
OY	1296	TACGACGCTGATCCGCTGTGCTGATTTTCCAGACTTCCGAGACCTGTCCACCTCTCAT	1355
Db	1265	TTTCACCCCTGATCCGCTGTGCTGATTTTCCAGACTTCCGAGACCTGTCCACCTCTCAT	1322
OY	1356	CCATGGGGCGAGGCTGTGATGTACATGACATCGGCTCCCTATTTTGGCGCATGG	1415
Db	1325	TCATGGGTGCGGAAGCTGCTGTGATGTGATGATGATGATGATGATGATGATGATGATGAT	1384
OY	1416	GAGCATTCAGCTCTTCCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1475
Db	1385	GGCCAAAGAGCTCTTCCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1444
OY	1476	CCCTTCAAGCTCATTTGGATGTATCTCCAAATGTTACTGAGAGAGGCAATGCTTTA	1535

Db	1445	TCCTTTCAATGTCATCTCTGGATGTCGTCTCCAAATGTCACCTCGAGAGGTCATAGCTGTGA	1504
QY	1536	CTATGAACGTGAAGAAGAGGGGCTGTAGACACACTGGTCCATATTTCTTTGGCAAGATCTCGG	1599
Db	1505	CTATGAGCTGGAAAGAGGGGCTGTAGACCTGCTGTCTCTTATTTCTTTGGCAGATCTTAGG	1566
QY	1596	GGAGCTTCGGGAGCACTGTGCTTACATCATCTACGGGATGCCACCCTACTGGCTGGC	1655
Db	1565	AGAAATGGCCGGAGCACTGTGCCCTAGGTCACTACACGGAGGCCCATCTACTGGCTGAC	1624
QY	1656	CAACCTGAGGCCAGGGGCTCCACCCCTCCCTGGTGCACCTCCGCTGGGTGTGGTGGTGT	1715
Db	1625	AAACCTGGCGGCCCGTGGCTGTAGAGCTCTTCTCTTACACTTCCGCTGGTGGTGGTGGT	1684
QY	1716	CTTCTGTTGACAGATTTATGGCCCTGGCCGCGCCGCGCCCTGGCTCCACCTCCACATAGGC	1775
Db	1685	CTTCTGTGTGACAGACCATATGGCCCTGGCTGGCTGTGCATGCTGCCACCTTCCACATATC	1744
QY	1776	CTCTCTTCTTACCAATGCCCTCTACACTCTTCTTACCTCGCCGGGGGCTTCATGATATA	1835
Db	1745	CTCTCTTCTTCTCAATGGCCCTCTACACCTCTTACCTTACTGCTGCGGCTTCATGATATA	1804
QY	1836	CTTGGAGCACCTGTGGAGACAGTCCCGCGCTGATTTTCCAAAGTGTCTTCCCTGGGTGTG	1895
Db	1805	CTTGGACACACCTGTGGATATGTCCCTGCATGATGTCCTCAAAAGTGTCTTCCCTGGGTGTG	1864
QY	1896	TTTGTGAAGGGCTGATGAAGATTTCACTTACAGCAGAGAACAATTATATAATGGCTCCGGGAA	1955
Db	1865	CTTCTGCGGGGCTGTATGACGATTTCAATTTAATGACACCTTTACACACACAATGGGAA	1924
QY	1956	CCTCACACATCGCGGTCTCAGAGATATAAATCCTCAGTCCATGTGAGAGCTGTGACTGCC	2015
Db	1925	CTTTCACCTTCTTCATCTCCTGGAGACACAGATGATAGTCCATGTGAGACCTGAACTGCCATCC	1984
QY	2016	TCCTACAGGCAATCTACCTCCATCGTCACTTGGGCTCAGGGGTGGCTTCATGCTCTGTACTA	2075
Db	1985	ACTGTATCGATCTACCTCATCTGATTTGTCATCGGATAGCTACGGCTGTCTCTCTGTACTA	2044
QY	2076	CGTGTCTTAAAGTTCATCAACAAAGAAACCAAGTCAAGACAGTGGTATTCACGGCCAGAGCT	2135
Db	2045	TCATATCTTGAAGCTCATCAACAAAGAAAGTCAATTCAAGACTGGTGTATCTAGCTTGTCT	2104
QY	2136	CTGCCCGCTGTGGGGGACCTGACAGACCCCTTCACTGCACTCCCTCCAGAGGCC	2195
Db	2105	CTCATCTGGCGG-----GACCCTTTTCCCGGGGGCTGGCCACCCAGAGGAGGCC	2152
QY	2196	TTTCTGGGGGACAGTGAAGACAATACCTTACAGATGCTCAAGCTACATCCGCGCCAGGGTG	2255
Db	2153	GGAGCTGGGGAAGAGGCTCACACAAATCTCTCAG-----GCAGCACCCACCTCTTGTGTG	2205
QY	2256	CTGCACTGGCACAGACCCAGCCAGAGGATGGCAGTAAATAAAGCACTGCCAAAGGATTT	2315
Db	2206	CTGCACTGGCACAGTCAAGCCACAGATGGCAGTAAATAAAGCACTGGAGGTGTGTT	2265
QY	2316	CTGCTCACTGCGAGGAGACTGCGATGACTGGGAGAAAACCTGCACTCGTGGCACTTACA	2375
Db	2266	CTGCTCCACGAGCCAGGCTGTGTATGGGAGAGAGAAA-----ACCAAGT	2310
QY	2376	ACGTTGCTAAATTTATTTCTTTTATATATGATTTATATAGCACTCGATATAGATGG	2435
Db	2311	ACGTTGCTCAATGCAATTT-----TATATCTTTAAATAAACAACCAAGTATAGATGGG	2362
QY	2436	AGCAAACTAGAAAGAATGGGTGTAGACTGCGAGGAATTTGGAAACCTGGAGGGA	2495
Db	2363	AAACAATTTATATAGATTTAGTATGCTATAGCGTATGCAAAATTTCTGGAAATCTGTAGAG	2422
QY	2496	ACAATTAACAGTAGACAGATTTGGCTCATCTTCCAGGGGCCCAACACTCGTGGTGA	2555
Db	2423	ATAGTGGTTTATACCAAGTGTATTAACCTTCTCTCTACATTTCTCAGC-----TGTATA	2478
QY	2556	GGCAACATCAATACAGAAATGACCTTAAGTATACACGACAAGTGC-CCATCCCTCTTTT	2614
Db	2479	GGCACTCCCATATCAAAAGGCGACATAAACAACACTAGCAAAATGTTTTTCTGTATCTTC	2538

OY	2615	TGTGAGGGTCATGGCGCTCAAAAGCCAACT	2646
Dd	2539	TGC GTGACTTCA TGAC TCC AACC CCAA GCT	2570
RESULT 2			
AKO50938			
LOCUS			
DEFINITION	AKO50938	2417 bp	mRNA linear HTC 05-DEC-2002
	Mus musculus	9 days embryo whole body cDNA, RIKEN full-length	
	enriched library, clone: D03040P06 product: ATP-BINDING CASSETTE,		
	SUB-FAMILY G, MEMBER 8 (STEROLIN-2) homolog [Mus musculus], full		
	insert sequence.		
VERSION	AKO50938		
KEYWORDS	AKO50938.1 GI:26094211		
SOURCE	HTC; CAP trapper.		
ORGANISM	Mus musculus (house mouse)		
REFERENCE	Eumayr et al., Metazoa: Chordata; Craniata: Vertebrata: Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	Carninci, P. and Hayashizaki, Y.		
JOURNAL	High-efficiency full-length cDNA cloning		
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)		
PUBMED	99279253		
REFERENCE	10349636		
AUTHORS	2		
TITLE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,		
JOURNAL	Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.		
MEDLINE	Normalization and subtraction of cap-trapper-selected cDNAs to		
PUBMED	prepare full-length cDNA libraries for rapid discovery of new genes		
REFERENCE	Genome Res. 10 (10), 1617-1630 (2000)		
AUTHORS	3		
TITLE	Shibata, K., Itoh, M., Alzawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,		
JOURNAL	Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M.,		
MEDLINE	Sumi, N., Ishii, I., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,		
PUBMED	Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,		
REFERENCE	Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, K.,		
AUTHORS	Taneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,		
TITLE	Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.		
JOURNAL	RIKEN Integrated sequence analysis (RISA) system--384-format		
MEDLINE	sequencing pipeline with 384 multiplexed sequencer		
PUBMED	Genome Res. 10 (11), 1757-1771 (2000)		
REFERENCE	20530913		
AUTHORS	11076861		
TITLE	4		
JOURNAL	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,		
MEDLINE	Alzawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuoka, S.,		
PUBMED	Arakawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondou, S., Yamana, I.,		
REFERENCE	Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,		
AUTHORS	Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Cassavant, T.,		
TITLE	Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochwa, H.,		
JOURNAL	Kuehl, P., Lewis, S., Matsuo, Y., Stabaldi, F., Suzuki, R., Tomita, M.,		
MEDLINE	Okazaki, Y., Schriml, L.M., Neuball, F., Suzuki, R., Tomita, M.,		
PUBMED	Wagner, L., Mashio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,		
REFERENCE	Mader, L., Mashio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,		
AUTHORS	Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bull, C.,		
TITLE	Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,		
JOURNAL	Hottelmer, C., Humé, D.A., Kamiya, M., Lee, N.H., Lyons, P.,		
MEDLINE	Marchionni, L., Mashima, J., Mazzarelli, J., Mommaerts, P., Nordone, P.,		
PUBMED	Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,		
REFERENCE	Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,		
AUTHORS	Toyooka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L.,		
TITLE	Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S.		
JOURNAL	and Hayashizaki, Y.		
MEDLINE	Functional annotation of a full-length mouse cDNA collection		
PUBMED	Nature 409 (6821), 685-690 (2001)		
REFERENCE	21085660		
AUTHORS	11217851		
TITLE	5		
JOURNAL	The FANTOM Consortium and the RIKEN Genome Exploration Research		
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			

TITLE
Group Phase I & II Team.
Analyses of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL
6 (bases 1 to 2417)

REFERENCE
Adachi,S., Alizawa,K., Akimura,T., Arikawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanaoka,T., Hara,A., Hashizume,M.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T.,
Horii,F., Imocani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Katoch,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohmoto,N.,
Okazaki,Y., Saito,R., Saitoh,H., Saitoh,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shingawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akihira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.

TITLE
Submitted (16-Jul-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration and Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama
Kanagawa 230-0045, Japan (E-mail:genome-res@gscc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)

COMMENT
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/
Location/Qualifiers

FEATURES
1. 2417
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM:DB:D030040P06"
/db_xref="taxon:10090"
/clone="D030040P06"
/tissue_type="whole body"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="9 days embryo"
1. 2417
/note="ATP-BINDING CASSETTE, SUB-FAMILY G, MEMBER 8
(STROLIN-2) homolog [Mus musculus] (SWISSPROT I09DBM0,
evidence: FASRY, 92%ID, 96.7%length, match-1796)"

BASE COUNT
554 a 695 c 603 g 565 t

ORIGIN

Query Match 48.2%; Score 1286.2; DB 11; Length 2417;
Best Local Similarity 76.1%; Pred. No.1.3e-308;
Matches 1696; Conservative 0; Mismatches 483; Indels 50; Gaps 7;

419 CAGGTTGGGAGAGGCTCTTCTTGAATGTGATCACTGGCCGAGAGTACAGGCCGACAGA 478
|||||
184 CAGCGTGGGAGAGGCTCTTCTTGAATGTGATCACTGGCCGAGAGTACAGGCCGAGAGTAC 243
|||||
479 TCAAGTCAGGCGACGATGTCGATTCATATGAGGACAGCCAGCTCGCTCAGCTGGGAGAGT 538
|||||
Db 244 TGAATATCAGCAAAATTTGGATTAATGAGGCAACCCAGTACGCTTCACTGGTGTGAGAACT 303
|||||
539 GTGTGGCCCGACGTGGCGCCAGGACCAACAGCTGTCTCCCACTGTGACTGTGGAGAGACT 598
|||||
Db 304 GCGTTGGCGCATGTGCGGACGATGACCAACTGCTGGCCCACTGACGCTCAGAGAGACCC 363
|||||
599 TGGGCTTCATGAGCCAGATGGGCGTGGCCGACAGACCTTCTCCAGAGCCAGAGGTGACAAA 658
|||||
Db 364 TGGCTTCATGAGCCAGATGGGCGTGGCCGACAGACCTTCTCCAGAGCCAGAGGTGACAAA 423
|||||
659 GGGTGGAGAGAGCTGATCGCGGAGCTGCGGCTTAAAGCACTGGCTGATGACACCGCGAGGCA 718
|||||

Db 424 GGGTGAAGACGTAATCGCCAGCTGCGGCTGCGGAGTGCACCAACAGAGTGGCA 483
OY 719 ACATGATACGTTGGGGGTTGTGGGGGTGTAGGCGAGAGATCAGATTTGGGTGAGC 778
Db 484 ACACGATGATGCTGGGGGTGTCCGGGGGTGAGCCGACAGATGAGATTTGGGTGAGC 543
OY 779 TCTGTGGAACCCAGGAAATCTTATTTCTGACGAAACCCAGCTTGGGCTGAGAGCTTCA 838
Db 544 TCTGTGGAACCCAGGAAATCTTATTTCTGATATACCCACTTGTGGCTTGCAGAGCTTCA 603
OY 839 CAGCCCAACACTGTGTGAGAACCTTGTCAAGGCTGCGCAAGGCAACCGGCTGTGTCTCA 898
Db 604 CAGCCCAACACTGTGTGAGAACCTTGTCCCGCTGCGCAAGGCAACAGGCTGTGTCTCA 663
OY 899 TCTCCCTCCAGCCGCTGCTGTGACATCTTCAGGCTGTTGATGTGTGCTCTCTGATGA 958
Db 664 TCTCCCTCCAGCCGCTGCTGTGACATCTTCAGGCTGTTGATGTGTGCTCTCTGATGA 723
OY 959 CGTCTGACACCCCTATCTATTAGGGGCGGCGGACAGATGTCAGATTTTCAACGCCA 1018
Db 724 CAGTGTGACACCCCTATCTATTAGGGGCGGCGGACAGATGTCAGATTTTCAACGCCA 783
OY 1019 TGGGTAACCCCTGTCTGTCTACAGAAATCTGCTGATCTTATGTGTGACCTGACAGCA 1078
Db 784 TGGGTAACCCCTGTCTGTCTACAGAAATCTGCTGATCTTATGTGTGACCTGACAGCA 843
OY 1079 TTGACGGGCGGACAGAGAGAGAAATTTGGCCACGAGAGAAAGCTCACTGCTGACAG 1138
Db 844 TGGACAGAGAGAGAGAGAGAAATTTGGCCACGAGAGAAAGCTCACTGCTGACAG 903
OY 1139 CCCTGTTTGTAGAAAAGTGTGCTGATGATGATCTTCTATGAAAGAGAGAGAGAGAG 1198
Db 904 CCCTGTTTGTAGAAAAGTGTGATGATGATCTTCTGTGTGAAAGCTGAGAGAGAG 963
OY 1199 ATCTTGAG 1258
Db 964 AACTCAACAG 1020
OY 1259 CGAGTGTCTGAGAAAG 1318
Db 1021 GAGAGTGTCTGAG 1080
OY 1319 TTTCACAG 1378
Db 1081 TTTCACAG 1140
OY 1379 TGTCAATGAG 1438
Db 1141 TGTCCCTCAATGAG 1200
OY 1439 ATACAG 1498
Db 1201 ATACAG 1260
OY 1499 TCAATCTCAATGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1558
Db 1261 TGTCTCCAAATGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
OY 1559 ACAG 1618
Db 1321 ACAG 1380
OY 1619 ACATATGATCTACAG 1678
Db 1381 AGTATATATCTACAG 1440
OY 1679 CTTTCTGTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1738
Db 1441 TCTTCTGTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
OY 1739 TGGCGCGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1798
Db 1501 TGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560

OY 1799 ACAATCTCTTCTACCTGCGCGGGGCTTCAATGATTAACCTTGACAGAGCTGTGACAGTGC 1858
Db 1561 ACAATCTCTTCTACCTTACTGCGGGCTTCAATGATTAACCTTGACAGAGCTGTGATGAGTGC 1620
OY 1859 CCGCGGAGATTTCCAAAGTGTCTCTGCGGGGTGTGAGAGGCTGTATGAAATTC 1918
Db 1621 CTGATGAGATCTCAAGCTGTCTCTGCGGGGTGTGAGAGGCTGTATGAAATTC 1680
OY 1919 AGTTACAG 1978
Db 1681 AATTATGAG 1740
OY 1979 ATAAATCTGATGCTGAG 2038
Db 1741 ACAGAGATGATGAG 1800
OY 2039 TCAATGCGCTGACGCGGTGCTGATGCTGTGATGATGATGATGATGATGATGATGATG 2098
Db 1801 TCAATGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1860
OY 2099 AGAAACCAAGTCAAG 2158
Db 1861 AGAAACCAAGTCAAG 1908
OY 2159 GCAAGACCTTCAACTGCT 2218
Db 1909 GACCTTCTTCCCGGGCTGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1968
OY 2219 GACCTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2278
Db 1969 GATCTTCTGAG-----GAG 2021
OY 2279 AGATGAG 2338
Db 2022 AGATGAG 2081
OY 2339 ATGATGAG 2398
Db 2082 ATGAG 2120
OY 2399 GATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2458
Db 2121 --TATATCTTAAATTAACAG 2178
OY 2459 AGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2518
Db 2179 AGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2238
OY 2519 TGGCTTCAATCTTCCAG 2578
Db 2239 TAACTTCTTCTTACATCTGACAGC-----TGTAAAGCAGCTCCCAATTAAGAGAGAG 2294
OY 2579 CCTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2637
Db 2295 CCTAAG 2354
OY 2638 AGCAACGT 2646
Db 2355 CCGCAAGT 2363

RESULT 3
HSM072006
ID HSM072006 standard; RNA; EST; 691 BP.
XX
AC BX481838;
SV BX481838.1
XX
DT 09-MAY-2003 (Rel. 75, Created)
XX 09-MAY-2003 (Rel. 75, Last updated, Version 1)
XX

DE	Homo sapiens mRNA; EST DKFZp686M06227_r1 (from clone DKFZp686M06227)
KM	EST: expressed sequence tag.
XX	
XX	
OS	Homo sapiens (human)
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia
CC	Eutheria; Primates; Catarrhini; Hominoidea; Homo.
XX	[1]
RP	1-691
RA	Bahr A., Lauber J., Mewes H.W., Well B., Amld C., Osanger A., Fobo G.,
RL	Han M., Wilemann S.;
XX	; Submitted (07-MAY-2003) to the EMBL/GenBank/DDBJ databases.
CC	MRS, Ingolstaedter Landstr.1, D-85764 Neuberg, GERMANY
CC	This is the 5' sequence of the clone insert
CC	Clone from S. Wilemann, Molecular Genome Analysis, German Cancer
CC	Research Center (DKFZ); Email s.wilemann@dkfz-heidelberg.de;
CC	sequenced by Oligen (Hilden/Germany) within the cDNA sequencing
CC	consortium of the German Genome Project.
CC	No sl sequence available.
CC	This clone (DKFZp686M06227) is available at the RZPD in Berlin.
CC	Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,
CC	14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX	
FH	Key Location/Qualifiers
FT	
FT	source 1..691
FT	/db_xref="taxon:9606"
FT	/mol_type="mRNA"
FT	/organism="Homo sapiens"
FT	/clone="DKFZp686M06227"
FT	/clone_1bp="686 (synonym: hlcc3). Vector psportl_sfl host
FT	DH10B; sites SfiI + SfiIB"
FT	/dev_stage="adult"
FT	/tissue_type="CDNA-collection"
XX	
SO	Sequence 691 bp; 161 A; 200 C; 168 G; 162 T; 0 other:
OY	Query Match 25.5%; Score 681.8; DB 2; Length 691;
OY	Best Local Similarity 99.7%; Prid. No. 1.8e-158;
OY	Matches 683; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY	1742 CCGCGGCGGGCCTGCCTGCCACCTCCACATCATGTAGCAAGTGGCGTCTTCACGAATGCCCTTACA 1801
OY	
OY	7 CGGCGACGAGGCGTGCTGCCACTCCACATCGGCCTCTTCACGATGCCCTTACA 66
OY	1802 ACTGCTTTACCTGCGCGGGGCTTCATGATAAATTGAGCACCTGTGGACAGTCCCG 1861
OY	
OY	67 ACTGCTTTACCTGCGCGGGGCTTCATGATAAATTGAGCACCTGTGGACAGTCCCG 126
OY	1862 CGTAGATTTCGAAGAAGTCTCTCTCGCGGGTCTTTGAAGGCTGATGAAATTCAGT 1921
OY	
OY	127 CGTAGATTTCGAAGAAGTCTCTCTCGCGGGTCTTTGAAGGCTGATGAAATTCAGT 186
OY	1922 TCAGCAGAAAGACTTTAAATAGCTCTCGGGAACCTCAACATCGGGTCTCAGAGATA 1981
OY	
OY	187 TCAGCAGAAAGACTTTAAATAGCTCTCGGGAACCTCAACATCGGGTCTCAGAGATA 246
OY	1982 AAATCTCTAGTCCATGAGAGCTGGACTCGTACCCTCTCTACGCGCATCTACCTATGTA 2041
OY	
OY	247 AAATCTCTAGTCCATGAGAGCTGGACTCGTACCCTCTCTACGCGCATCTACCTATGTA 306
OY	2042 TTGGCCCTCAGCGGTGGCTTCATGTCCTGCTACAGCTGTCCTTAAGTTCATCAACAGA 2101
OY	
OY	307 TTGGCCCTCAGCGGTGGCTTCATGTCCTGCTACAGCTGTCCTTAAGTTCATCAACAGA 366
OY	2102 AACCAAGTCAAGACTGTGATTCACGCCAGAGCTGCCCCGTGTGTGGGGAGCTGAGGA 2161
OY	
OY	367 AACCAAGTCAAGACTGTGATTCACGCCAGAGAGTGTGCCCGTGTGTGGGGAGACTGAGGA 426
OY	2162 GACCCTTCAACTGACTCCCTCTCTCAGGAGCCCTTCTGCGGAGACATGAGAACATGAC 2221

Db	427	GACCCTCAACTGCACACTCCCTCTCAGSAGGCCCTTCTCTGGGACAGTGAAGACAATATAC	486
QY	2222	CCTACAGATGCTCAGCTATACCGGCCCAAGGCTCTGCATGTGCACACAGCACCACAGG	2287
Db	487	CCTACAGATGCTCAGCTACATCCGGCCCAAGGCTCTGCATGTGCACACAGCACCACAGG	546
QY	2282	ATGSCAGTAGAATAAAGACACTGCAGAAAGGATTTCTGCTCACTGGCAGAGACTGCGATG	2341
Db	547	ATGSCACTAGAAATAAAGACACTGCAGAAAGGATTTCTGCTCACTGGCAGAGACTGCGATG	606
QY	2342	ACTGSGAGAAACCTGCACCTCGGGTGGCACTACACAGCTTGTAATTTATTTCCCTTTGAT	2403
Db	607	ACTGSGAGAAACCTGCACCTCGGGTGGCACTACACAGCTTGTAATTTATTTCCCTTTGAT	666
QY	2402	ATGCATTTATTTAGGCAACTCGAATA	2426
Db	667	ATGCATTTATTTAGGCAACTCGAATA	691

[illegible]

FEATURES	location/Qualifiers
source	l. 849
	/organism="Mus musculus"
	/mol_type="mRNA"
	/strain="FVB/N"
	/db_xref="taxon:10090"
	/clone_image="5135115"
	/lab_host="DH10B (T1 phage-resistant)"
	/clone_lib="NCI CGAP L19"
	/note="Organ: liver; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI CGAP Library."
BASE COUNT	179 a 251 c 203 g 216 t
ORIGIN	
Query Match	17.3%; Score 460.6; DB 12; Length 849;
Best Local Similarity	77.3%; Pred. No. 2.2e-103;
Matches 639; Conservative	0; Mismatches 174; Indels 14; Gaps 6;
QY	991 CAGCACATGTCAGTATTTCACAGCAGTCGCTACCCCTGCTCGCTACAGCAATCCCT 1050
DB	1 CAGCAATGTCGACGTAACCTTCACATCATCATTCATGGCCACCCCTTGTCTCGCTATACCAACCT 61
OY	1051 GCTGACTTCTATGTGACACCTGCACAGCATTTGACAGCGCAGCAGACAGAGCAATTTGCC 1110

Db	62	CGGGACTTCTAGSTGACTTGACCAAGCATTCAGACAGACGACACCAAGAAAGGAGGTGGCC	121	
OY	1111	ACGAGGGAGAAAGGCTCACTCACTGCGAGCCCTGTTCTAGAAAAAGTGGCTGACTTAGAT	11707	
Db	1122	ACCGTGGAGAAAGGCAAGCTCTCTTGACACCCGTTGCTTAGAAAAAGTACAAAGGCTTTGAT	181	
OY	1171	GACTTTCATGGAAGGAGAGACGAGAGATCTTGACGAGGACACTGTGTGGAAAGCAGC	1230	
Db	182	GACTTTCGTGGAAAGCTGAGGCAAGAGACTCAACACAGACGCCACACAGTGAAGCTG	241	
OY	1231	GTGACCCCACTAGACACCACTGCTCCGAGTGCCTTAGAGATGCTGGGGGCGGTGAC	1290	
Db	242	ACCCCTACACAGAGACACTGACTG---TGGAGCTGCTGTTAGAGCTGCCGGATGATAG	298	
OY	1231	CAGTTTACGAGCCCTGATCCGTGTCAGATTTTCCACAGCATTCGAGACCTGCCACCTC	1350	
Db	299	CAGTTTCCACCCCTGATCCGTGTCAGATTTTCCAAATGACTCCGGAGACCTGCCACGCTG	358	
OY	1351	CTCATCCATGAGGGGGGAGGCGCTGTCTGATATGTCATATGACATCAGGCTTCTTATTTGGC	1410	
Db	359	CTCATATTCAGGGGTGGAGAGCTGCTGATGTCCTCATATTCATATGCTTCTTACTAGGCG	418	
OY	1411	CATGGGAGCATTCAGACTCTCCCTTCATGATATGATACAGCCGCGCTTGTCATGATGCTGCT	1470	
Db	419	CATGGGGGCAAGCAGCTCTCTCTTCATGAGACAGACAGACGCCCTCTTCATATAGGGGCG	478	
OY	1471	CTCATCCCTTTCAACAGCTCATATCTGAGTCTATCTCCAAATGTTACTAGAGAGGGCATG	1530	
Db	479	CTCATTCCTTTCAATATGTCATCTGAGATGCTGTCCAAATGTCACTCGGAGAGGTCAATG	538	
OY	1531	CTTATAGTATGAAACGGAAGACGGGCTGTACACCACTGGTCCATATTTCTTGGCAATATC	1590	
Db	539	CTGTACTATGAGCTGGAAGACGGGCTGT---ACTGCTGTCTCTATTTCTTGGCAATATC	596	
OY	1591	CTCGGGAGGCTTCCGAGGACAC-TGTGCTATCATATCATATGATACGGAGATGCCACCTACTG	1649	
Db	597	CTAGAGAAATGCGCGAGCACTTGTGCTGATGATCATATCATGCGCATGCCCATCTACTG	656	
OY	1650	GCTGGCCCACTGAGGCCAGCCAGGCTCCAGACCCCTTCCTGCTGCTG---ACTTCTGCTGGTGG	1707	
Db	657	GCTGACAAACCTGCGGGCCGTGCTGAGCTTCTTCTTCAACACTTTCCTGCTGGTGG	716	
OY	1708	CTGGTGTCTTCTGCTGAGAGTATATGCCCCCTGGCCCGCGGCGCTGCTCCCACTTC	1767	
Db	717	TAGAGAGTCTTCTGCTGAGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	772	
OY	1768	CACATGCGCTCTCTCTTACAGCAATGCCCTTACACACTCTTACTT	1814	
Db	773	CACATGCTCTCTCTCTGCA--TGCCTTGAATCTTCTTACTT	817	
RESULT 5	BP660076	549 bp	linear	EST 20-DEC-2000
LOCUS	BP660076	maaz7c08.y1	NCI CGAP L110	Mus musculus cDNA clone IMAGE:3812342 5
DEFINITION		similar to TR:Q9V0N4 Q9V0N4 G99664	PROTEIN. ; mRNA sequence.	
ACCESSION	BF660076			
VERSION	BF660076.1	GI:11925210		
KEYWORDS		EST.		
SOURCE		Mus musculus (house mouse)		
ORGANISM		Mus musculus (house mouse)		
REFERENCE				
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus; 1 (bases 1 to 549)		
TITLE		NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .		
JOURNAL		National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
COMMENT		Unpublished		
		Other_ESTs: maaz7c08.x1		
		Contact: Robert Strausberg, Ph.D.		
		Email: cgaps-remail.nih.gov		
		Tissue Procurement: Jeffrey E. Green, M.D.		
		cDNA Library Preparation: Life Technologies, Inc.		

[illegible]

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
BY705076	583 bp mRNA linear	EST 16-DEC-2002						
BY705076	RIKEN full-length enriched, adult male liver	Mus musculus						
BY705076	CDNA clone 1300003C16 5', mRNA sequence.							
BY705076	GI:27116215							
EST								
Mus musculus								
Mus musculus								
Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.								
1 (bases 1 to 583)								
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Mikado, I., Otsu, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gotohori, T., Baldarelli, R., Hill, D. P., Bul, C., Hume, D. A., Quackenbush, J., Schmitt, L. M., Knapman, A., Matsuda, H., Batalov, S., Betsel, K. W., Blake, J. A., Brad, D., Brusic, V., Chohts, C., Corbani, L. E., Coustas, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, D., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, J. J., Jarvis, E. D., Kana, A., Kawai, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltas, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, F., Pavan, W. J., Pereira, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sanderlin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, R., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wyshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavalina, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Maki, K., Kawai, I., Aizawa, K., Aikawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shiga, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E., and Hayashizaki, Y.								
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs								
Nature 420, 563-573 (2002)								
22354683								
12466851								
CONTACT: Yoshinori Hayashizaki								
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute								
The Institute of Physical and Chemical Research (RIKEN)								
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan								
Tel: 81-45-503-9222								
Fax: 81-45-503-9216								
Email: genome-resgsc.riken.go.jp, URL: http://genome-gsc.riken.go.jp/								
Adachi, J., Aizawa, K., Akimura, T., Aikawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Koye, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takada, Y., Maki, K., Metahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission								
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)								
Normalization and subtraction of cap-trapper-selected cDNAs to Prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)								
RIKEN Integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)								
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)								
CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.								

Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES
source
Location/Qualifiers
1..583
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1300003C16"
/sex="male"
/tissue_type="liver"
/dev_stage="adult"
/clone_lib="RIKEN full-length enriched, adult male liver"

BASE COUNT
ORIGIN
159 a 149 c 169 g 106 t

Query Match
Best Local Similarity 82.7%; Pred. No. 8,8e-79;
Matches 426; Conservative 0; Mismatches 86; Indels 3; Gaps 1

OY 99 CATGGCCGGGAAGGCGGCAGAGAGAGAGGGCTGCCGAAGGGGCACTCCCAGATAC 158
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68 CATGGCTAGAANAACAAGAAAGAGACCACCTGTGAATGAGACTGTACTTCAGATGC 127
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 159 CTC---GGGCGCTCCAGGATATATGTCTTCTCTGAAGTGACACACCTGTACTTCAC 215
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 128 TTCCAGAGGCTCCACGACACTGTCTTCTCTCGAAGTAGAACACACTGTACTTCAC 187
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 216 CTACAGTGGCCAGCCCAACACCCTGTGAGGTGAGAGACTCAACTACAGTGGACCTGGC 275
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 CTACAGTGTGATGCCAACACTCTGTGAGGTGAGAGATTCTCATCTACAGTGGACATGCG 247
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 276 CTCCTCAGGTCCCTTGTTGAGACAGCTGGCTCAGTTCAAAGATGCCCTGGACATCTCCAG 335
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 248 CTCCTCAGGTGCTTGTGTTGAGACAGCTGGCTCAGTTCAAAGATGCCCTGGACATCTCATAG 307
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 336 CTGCCAGAAATCTTGTGAGCTGGGCAATCCAGAACCTTAAGCTTCAAGTAGAGAGAGGGCA 395
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 308 CAGCCAAAGACTCGTGTGAGCTGGGGCATGCCAAATCTAAGCTTCAAGTAGAGAGAGGACA 367
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 396 GATCTGCGCCATCATAGGAGAGCTCAGGTTGTGGAGAGACCTCTCTTCTAGATGTGATCAC 455
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 368 GATCTGCGCCATCATAGGAGAGCTCAGGTTGTGGAGAGACCTCTCTTCTAGATGTGATCAC 427
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 456 TGGCCGAGGTACAGCGCGCAAGTCAAGTCAAGGCGAGATCGAATGGAATGGGCGCCAG 515
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 428 AGGCAAGAGCCACAGCTGGCAAGATGAATTCAGCAAAATTTGATTAATGGGCAACCCAG 487
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 516 CTCGCTCAGCTGTGTGAGGAAGTGTGTGGCCACAGTGGGCCAGCAACACAGCTCTCC 575
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 488 TACCCCTCAGCTGTGTGAGGAAGTGTGTGGCCAGTGTGGGAGCATGACCAACATGCTCC 547
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 576 CAACTTGACTGTGGAGAGACCTTGGGCTTCAATG 610
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 548 CAACCTGAACCTCAGAGAGACCTGTGCTTCATTG 582
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
T91380/c
LOCUS
DEFINITION yd3b02.s1 Soares fetal liver spleen INFLS Homo sapiens CDNA clone
IMAGE:111915 3', mRNA sequence.
T91380
ACCESSION T91380
VERSION T91380.1 GI:721293
KEYWORDS EST
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 457)
Hillier,L., Clark,N., Dubque,T., Elliston,K., Hawkins,M., Holtman

OY 1471 CTCATCCCTTCAACGCTATTCGTGATGTCATCTCCAAATGTACTCAGAGGCGCAATG 1530
 DB 181 CTCATCCCTTCAACGCTATTCGTGATGTCATCTCCAAATGTACTCAGAGGCGCAATG 240
 OY 1531 CTTTACTATGAACCTGAGAGCGGCTGTACACACCTGCGCATATTTCTTGGCCCAAGATC 1590
 DB 241 CTTTACTATGAACCTGAGAGCGGCTGTACACACCTGCGCATATTTCTTGGCCCAAGATC 300
 OY 1591 CTCGGGAGCTTCGCGAGCAGCTGTGCTTACATCA 1624
 DB 301 CTCGGGAGCTTCGCGAGCAGCTGTGCTTACATCA 334
 RESULT 9
 LOCUS BB610072 510 bp mRNA linear EST 26-OCT-2001
 DEFINITION BB610072 RIKEN full-length enriched, adult male liver Mus musculus
 ACCESSION BB610072
 VERSION BB610072.1 GI:16451685
 KEYWORDS EST
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 510)
 REFERENCE
 AUTHORS Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,
 Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Kono, H., Kouda
 M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okada
 Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki
 D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,
 Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,
 Muramatsu, M. and Hayashizaki, Y.
 RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
 TITLE Unpublished
 JOURNAL Contact: Yoshinobu Hayashizaki
 COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Saitama-cho, Tsukuba, Ibaraki, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsr.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
 M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 Wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
 Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
 Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
 Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamane, I., Aizawa
 K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
 Hayashizaki, Y.
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.
 e mouse tissues.
 Location/Qualifiers
 1. 510
 /organism="Mus musculus"
 /mol_type="mRNA"

OY 99 CATGCCGCGAAGCGGCGAGAGAGAGGCGCTGCCCAAGGCGCCACTCCCAAGGATAC 158
 DB 63 CATGCCGCGAAGCGGCGAGAGAGAGGCGCTGCCCAAGGCGCCACTCCCAAGGATAC 122
 OY 159 CTCGGGCTCCAGAGATGATTTCTCTCTGAAAGTACAAAGCCCTGTACTTACCTA 218
 DB 123 TTCGGGCTCCAGAGAGAGGCGCTGCCCAAGGCGCCACTCCCAAGGATAC 182
 OY 219 CAGTGGCCAGCCCAACACCTGAGAGTCAAGACCTCACTACAGGTGAGCTGGCCTC 278
 DB 183 CAGTGGCCAGCCCAACACCTGAGAGTCAAGACCTCACTACAGGTGAGCTGGCCTC 242
 OY 279 TCAGTCCCTGCTTTGAGAGAGGCGCTGCTCAATGATGCTTCCAGATCTCCAGCTG 338
 DB 243 TCAGTCCCTGCTTTGAGAGAGGCGCTGCTCAATGATGCTTCCAGATCTCCAGCTG 302
 OY 339 CCAAGATTCCTGTAGAGTGGGATCCAGAACCTTAAGCTTCAAGTGAAGTGGAGAT 398
 DB 303 CCAAGATTCCTGTAGAGTGGGATCCAGAACCTTAAGCTTCAAGTGAAGTGGAGAT 362
 OY 399 GCTGCGCATCTATGAGAGCTGAGTGTGGAGAGCCCTCTTGTAGATGATCACTGG 458
 DB 363 GCTGCGCATCTATGAGAGCTGAGTGTGGAGAGCCCTCTTGTAGATGATCACTGG 422
 OY 459 CCGAGTCAAGAGGCGGCGAAGATCAAGTCAAGGCAATGATGATGAGGAGCCAGCTC 518
 DB 423 CAGAGGCGGCGGCGGCGAAGATCAAGTCAAGGCAATGATGATGAGGAGCCAGCTC 482
 OY 519 GCTGCGCATCTATGAGAGCTGAGTGTGGAGAGCCCTCTTGTAGATGATCACTGG 546
 DB 483 GCTGCGCATCTATGAGAGCTGAGTGTGGAGAGCCCTCTTGTAGATGATCACTGG 510
 RESULT 10
 LOCUS A1157365 511 bp mRNA linear EST 30-SEP-1998
 DEFINITION A1157365 Y1 Sugano mouse embryo mewa Mus musculus cDNA clone
 ACCESSION A1157365
 VERSION A1157365.1 GI:3685834
 KEYWORDS EST
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 511)
 REFERENCE
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Stepien, M., Tan, F., Underwood, K., Moore, B.,
 Pfeister, R., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 The WashU-HMI Mouse EST Project
 Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810

QY 99 CATGCCCGGGAAGGGGCACAGCAGAGAAGGCGCTGCCGAAAGGGCCACTCCCAGCATAC 158
 | | | | | | | | | | | | | | | | | | |
Dd 60 CATGCGCTGAGAAAACAAGAAGAACCCAGCTGTGGAAATGGGACTTCACTTCAGAGATTGC 119

```

/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone IMAGE:930988"
/tissue_type="diaphragm"
/dev stage="adult"
/lab host="SOLR (kanamycin resistant)"
/clone_idb="Stratagene mouse diaphragm (#937303)"
/notes="Organ: diaphragm; Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally from mRNA

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 27, 2003, 00:38:04 ; Search time 569.074 Seconds
(without arguments)

9675:628 Million cell updates/sec

Perfect score: 2669

Sequence: 1 ggtgccctgtctccaggaac.....caatlaaatgtattgagc 2669

scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Maximum DB seq length: 20000000000

```

post-processing: minimum match 0%
                  maximum match 100%

```

Database : Published_Applications_NA:*

- 1: /cgn2_6/prodata/2/pub/pna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/prodata/2/pub/pna/US07_NEW_PUB.seq.*
- 3: /cgn2_6/prodata/2/pub/pna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/prodata/2/pub/pna/US07_PUBCOMB.seq.*
- 5: /cgn2_6/prodata/2/pub/pna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/prodata/2/pub/pna/US08_PUBCOMB.seq.*
- 7: /cgn2_6/prodata/2/pub/pna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/prodata/2/pub/pna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/prodata/2/pub/pna/US09_PUBCOMB.seq.*
- 10: /cgn2_6/prodata/2/pub/pna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/prodata/2/pub/pna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/prodata/2/pub/pna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/prodata/2/pub/pna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/prodata/2/pub/pna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/prodata/2/pub/pna/US10_NEW_PUB.seq.*
- 16: /cgn2_6/prodata/2/pub/pna/US60_NEW_PUB.seq.*
- 17: /cgn2_6/prodata/2/pub/pna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2669	100.0	2669	11	US-09-989-981A-7	Sequence 7, Appl
2	1430	53.6	2619	11	US-09-989-981A-3	Sequence 3, Appl
3	203.6	7.6	2340	9	US-09-837-992-4	Sequence 4, Appl
4	193.6	7.6	2340	11	US-09-989-981A-5	Sequence 5, Appl
5	193.4	7.2	1959	11	US-09-989-981A-1	Sequence 1, Appl
6	193.4	7.2	2258	9	US-09-837-992-2	Sequence 2, Appl
7	122	4.6	2025	10	US-09-866-866A-13	Sequence 13, App
8	115.4	4.3	2247	10	US-09-866-866A-26	Sequence 26, App
9	115.4	4.3	2418	11	US-09-961-086-2	Sequence 2, Appl
10	115.4	4.3	2574	10	US-09-981-353-14	Sequence 34, App
11	115.4	4.3	2718	14	US-10-130-687-60	Sequence 60, App
12	115.4	4.0	2719	10	US-09-866-866A-9	Sequence 9, Appl
13	106.8	4.0	3201	13	US-10-072-621-5	Sequence 5, Appl
14	105.2	3.9	2930	10	US-09-954-531-591	Sequence 591, Appl
15	105.2	3.9	2930	14	US-10-171-581-276	Sequence 276, Appl
16	101.4	3.8	447	10	US-09-960-352-11649	Sequence 11649, Appl

C	17	93.6	3.5	427	10	US-09-960-352-12839	Sequence 12839, A
C	18	93.6	3.5	2400	13	US-10-108-605-244	Sequence 244, App
C	19	92.2	3.5	2788	9	US-09-745-763-196	Sequence 196, App
C	20	92	3.4	10330	13	US-10-001-189-68	Sequence 68, App
C	21	90.6	3.4	6043	11	US-09-989-981A-9	Sequence 9, App
C	22	90.2	3.4	377	10	US-09-960-552-4326	Sequence 4326, App
C	23	89.4	3.3	414	10	US-09-960-352-5514	Sequence 5514, App
C	24	88.2	3.3	2223	10	US-09-938-842A-2262	Sequence 2262, App
C	25	87.6	3.3	387	10	US-09-960-352-6470	Sequence 6470, App
C	26	87.2	3.3	1941	14	US-10-090-455-3	Sequence 3, App
C	27	87.2	3.3	2687	13	US-10-154-452-3	Sequence 3, App
C	28	87.2	3.3	2687	14	US-10-090-455-12	Sequence 12, App
C	29	87.2	3.3	3455	13	US-10-072-621-4	Sequence 4, App
C	30	87.2	3.3	3455	14	US-10-090-455-1	Sequence 1, App
C	31	86.8	3.3	2687	13	US-10-154-452-7	Sequence 7, App
C	32	83.2	3.1	2340	14	US-10-156-761-1148	Sequence 1148, App
C	33	83.2	3.1	9025608	14	US-10-083-357-662	Sequence 1, App
C	34	81.4	3.0	1635	14	US-09-764-877-2722	Sequence 662, App
C	35	76.2	2.9	986	10	US-09-294-093B-4811	Sequence 2722, App
C	36	61.2	2.3	210	9	US-09-960-352-9037	Sequence 4811, App
C	37	57.4	2.2	2529	14	US-10-156-761-6449	Sequence 6449, App
C	38	57.4	2.2	9025608	14	US-09-938-842A-1103	Sequence 1, App
C	39	56.8	2.1	2118	10	US-10-156-761-1	Sequence 1103, App
C	40	56.4	2.1	1221	14	US-10-156-761-3499	Sequence 3499, App
C	41	55.8	2.1	3376	14	US-10-037-270-918	Sequence 918, App
C	42	55.2	2.1	774	14	US-10-156-761-4059	Sequence 4059, App
C	43	54	2.0	1936	14	US-10-156-761-5662	Sequence 5662, App
C	44	51.6	1.9	354	10	US-09-960-352-9037	Sequence 9027, App
C	45	51.6	1.9	472	11	US-09-918-993-30637	Sequence 30637, A

ALIGNMENTS

```

RESULT 1
US-09-989-981A-7
Sequence 7, Application US/09989981A
Publication No. US20030049730A1
GENERAL INFORMATION:
APPLICANT: Hobbs, Helen H.
APPLICANT: Shan, Bei
APPLICANT: Barnes, Robert
APPLICANT: Tian, Hui
APPLICANT: Tularik Inc.
APPLICANT: Board of Regents, The University of Texas System
TITLE OF INVENTION: ABC5 and ABC8: Compositions and Methods of Use
FILE REFERENCE: 018781-00732005
CURRENT APPLICATION NUMBER: US/09/989,981A
CURRENT FILING DATE: 2002-07-23
PRIOR APPLICATION NUMBER: US 60/252,235
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/253,645
PRIOR FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 7
LENGTH: 2669
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (100)..(2121)
OTHER INFORMATION: human ABCG8 (hABCG8)
US-09-989-981A-7

Query Match          100.0%; Score 2669; DB 11; Length 2669;
Best Local Similarity 100.0%; Pred No. 0;
Matches 2669; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GTGTCCTGCTCTCAGGAACAGAGTAGTGAAGACACTGGCCCTGGCAGGCAGCAGCTGGGCT 60
|||||
1 GTGTCCTGCTCTCAGGAACAGAGTAGTGAAGACACTGGCCCTGGCAGGCAGCAGCTGGGCT 60
|||||

```

OY	61	AAGAGAGCTGCAGGCCAGGAGTTCACAGAGCTGTGGGCCCATATGCGGGAGAGCGCAGAG	120
Db	61	AAGAGAGCTGCAGGCCAGGAGTTCACAGAGCTGTGGGCCCATATGCGGGAGAGCGCAGAG	120
OY	121	GAGAGAGGCTGCCGAAAAGGGGGCCACTCCCCGAGATACCTGGGGCTCCAGGATGATTG	180
Db	121	GAGAGAGGCTGCCGAAAAGGGGGCCACTCCCCGAGATACCTGGGGCTCCAGGATGATTG	180
OY	181	TTCTCTCTGAAAGTGACAAACAGCCCTGTACTTCACCTACAGTGGCGACGCCAAACCTTG	240
Db	181	TTCTCTCTGAAAGTGACAAACAGCCCTGTACTTCACCTACAGTGGCGACGCCAAACCTTG	240
OY	241	GAGGTCAAGAGACTTCACCTACAGGTGGACCTGSCCTCTCAGGCTCTTGTGAGCAG	300
Db	241	GAGGTCAAGAGACTTCACCTACAGGTGGACCTGSCCTCTCAGGCTCTTGTGAGCAG	300
OY	301	CTGGGTGAGTTCAAGATGCGCTGGGCAATCTCCACCTGCAGAAATTTCTTGATAGCTGGGC	360
Db	301	CTGGGTGAGTTCAAGATGCGCTGGGCAATCTCCACCTGCAGAAATTTCTTGATAGCTGGGC	360
OY	361	ATCCGAAACCTAAGCTTCAAGATGAGAGTGGGCGAGATGCTGGCCATATAGAGAGCTCA	420
Db	361	ATCCGAAACCTAAGCTTCAAGATGAGAGTGGGCGAGATGCTGGCCATATAGAGAGCTCA	420
OY	421	GCTTGTGGGAGAGCCTCCTTGGTAGATGTATCACTGGGCCGAGTCAACGGCGCAAGATC	480
Db	421	GCTTGTGGGAGAGCCTCCTTGGTAGATGTATCACTGGGCCGAGTCAACGGCGCAAGATC	480
OY	481	AAGTCAGGCCAGATCTGGATCAATGGGCGAGCCCACTGCTCAGCTGTGTAGAGAAAGT	540
Db	481	AAGTCAGGCCAGATCTGGATCAATGGGCGAGCCCACTGCTCAGCTGTGTAGAGAAAGT	540
OY	541	GTGGCCCCACGTGCGCCACACACACAGCTGCTCCCACTTGACTGTGCSAGAGACCTTG	600
Db	541	GTGGCCCCACGTGCGCCACACACACAGCTGCTCCCACTTGACTGTGCSAGAGACCTTG	600
OY	601	GCTTCATATGCCCCAGATCGCGCTGCCCAACCTTCTCCAGGCCACGTCACAAAAG	660
Db	601	GCTTCATATGCCCCAGATCGCGCTGCCCAACCTTCTCCAGGCCACGTCACAAAAG	660
OY	661	GTGGAGGACGTATCGGGGAGCTGCGGCTTAGCGATGCGCTGACACCCGGTGGGCAAC	720
Db	661	GTGGAGGACGTATCGGGGAGCTGCGGCTTAGCGATGCGCTGACACCCGGTGGGCAAC	720
OY	721	ATGTACGTGCGGGGGTGTGCGGGGGGTAGCGCAGAGAGACATTTGGGGTGGCAGCTC	780
Db	721	ATGTACGTGCGGGGGTGTGCGGGGGGTAGCGCAGAGAGACATTTGGGGTGGCAGCTC	780
OY	781	CTGTGGAACCCAGGAATCTTATCTCGAGAAACCCACCTGTGGGCTCGACAGCTTCACA	840
Db	781	CTGTGGAACCCAGGAATCTTATCTCGAGAAACCCACCTGTGGGCTCGACAGCTTCACA	840
OY	841	GCCCCAACAACCGTGGAAGACCTGTGCGCAGGCTGGCCAAAGCAACCGGCTGTGCTATC	900
Db	841	GCCCCAACAACCGTGGAAGACCTGTGCGCAGGCTGGCCAAAGCAACCGGCTGTGCTATC	900
OY	901	TCCCTCCACACAGCCTGCTGTGACATTTCAAGGGCTTTGATCTGTGCTCTGATGACG	960
Db	901	TCCCTCCACACAGCCTGCTGTGACATTTCAAGGGCTTTGATCTGTGCTCTGATGACG	960
OY	961	TCTGGACATCCCATCTACTTATGAGGGGGCGCCAGCAATGTGTCAGATATTTACAAGCCATC	1020
Db	961	TCTGGACATCCCATCTACTTATGAGGGGGCGCCAGCAATGTGTCAGATATTTACAAGCCATC	1020
OY	1021	GGCTAACCCCTTCCCTCGCTACAGCAATCTCTGACTTCTATGTGAGACTTACACAGCATT	1080
Db	1021	GGCTAACCCCTTCCCTCGCTACAGCAATCTCTGACTTCTATGTGAGACTTACACAGCATT	1080
OY	1081	GACAGGGCGACAGAGAGACAGAAATTTGGCCACACAGGAGAGGCTCACTCGCAGCC	1140
Db	1081	GACAGGGCGACAGAGAGACAGAAATTTGGCCACACAGGAGAGGCTCACTCGCAGCC	1140
OY	1141	CTGTTTTCAGAAAAAGTGCCTGACTTGATATGACTTTCTATGAGAAACGACAGCAAGAT	1200

Db	1141	CTGTTTCTAGAAAAAGTGCGTACTTATGATGACTTTCTATGGAAGCAGACGAAAGAT	1200
OY	1201	CTTGACGAGGACACCTGTGTGGAAAAGCAGCGCTGACCCACTAGACACCACTGCTCCCG	1260
Db	1201	CTTGACGAGGACACCTGTGTGGAAAAGCAGCGCTGACCCACTAGACACCACTGCTCCCG	1260
OY	1261	AGTCTTAGAAGATGCTCTGGGGCGGTGAGAGAGTTTACGAGCGTATCCGTGCAGAT	1320
Db	1261	AGTCTTAGAAGATGCTCTGGGGCGGTGAGAGAGTTTACGAGCGTATCCGTGCAGAT	1320
OY	1321	TCCAACGACTTCCGAGACCTGCCACCCCTCTCATTCATGAGGGGCGAGGCCCTGTCTGATG	1380
Db	1321	TCCAACGACTTCCGAGACCTGCCACCCCTCTCATTCATGAGGGGCGAGGCCCTGTCTGATG	1380
OY	1381	TCATGACCATGCGGCTTCTATTTTGGCCATGGAGATCAGACCTCTCTCATGAT	1440
Db	1381	TCATGACCATGCGGCTTCTATTTTGGCCATGGAGATCAGACCTCTCTCATGAT	1440
OY	1441	ACAGCGCGCTCTTGTTCATGATGAGTGGTCTCTCATCCCTTCAACGTCAATTCTGGATGTC	1500
Db	1441	ACAGCGCGCTCTTGTTCATGATGAGTGGTCTCTCATCCCTTCAACGTCAATTCTGGATGTC	1500
OY	1501	ATCTCCAAATGTACTCAGAGAGGGCATGCTTTACTATGAACTGGAAGAGCGGCTGTAC	1560
Db	1501	ATCTCCAAATGTACTCAGAGAGGGCATGCTTTACTATGAACTGGAAGAGCGGCTGTAC	1560
OY	1561	ACCACTGGTCCATATTTCTTTGCCAAGATCTCTGGGAGCTTCCGAGCACTGTGCTTAC	1620
Db	1561	ACCACTGGTCCATATTTCTTTGCCAAGATCTCTGGGAGCTTCCGAGCACTGTGCTTAC	1620
OY	1621	ATCATCATCTACGAGGATGCCACACATAGCGGTGGGCAACCTGAGGCCAGGCGCTCCAGGCC	1680
Db	1621	ATCATCATCTACGAGGATGCCACACATAGCGGTGGGCAACCTGAGGCCAGGCGCTCCAGGCC	1680
OY	1681	TTCCGTGTCGACTTCTGCTGCTGCTGGTGGTGTGCTCTCTGTGTCAGAGATTATGACCCTG	1740
Db	1681	TTCCGTGTCGACTTCTGCTGCTGCTGGTGGTGTGCTCTCTGTGTCAGAGATTATGACCCTG	1740
OY	1741	GCGCGCGGCGCCCTGCTGCCACTTTCACATGAGCTCTCTTCAAGATGCCCTGTAC	1800
Db	1741	GCGCGCGGCGCCCTGCTGCCACTTTCACATGAGCTCTCTTCAAGATGCCCTGTAC	1800
OY	1801	AACCTCTTCTACCTCGCGCGGGGGCTTACATGATTAAGTGAAGCAGCTGTGGACAGTGGCC	1860
Db	1801	AACCTCTTCTACCTCGCGCGGGGGCTTACATGATTAAGTGAAGCAGCTGTGGACAGTGGCC	1860
OY	1861	GCGTGGATTTCCAAAGTGTCCCTCTCGCGGGTGTGTTTGAAGGGCTGATGAATTTACG	1920
Db	1861	GCGTGGATTTCCAAAGTGTCCCTCTCGCGGGTGTGTTTGAAGGGCTGATGAATTTACG	1920
OY	1921	TTCCAGCAGAAACTTTATAAAATGCTCTGGGAACTCAGCATGCGGGTCTCAGAGAT	1980
Db	1921	TTCCAGCAGAAACTTTATAAAATGCTCTGGGAACTCAGCATGCGGGTCTCAGAGAT	1980
OY	1981	AAATTCCTCATGTGCATGAGACCTGGACCTGACCTCTACAGGCATCTACCTATGTC	2040
Db	1981	AAATTCCTCATGTGCATGAGACCTGGACCTGACCTCTACAGGCATCTACCTATGTC	2040
OY	2041	ATTGGCCTCAGCGGTGCTTCATGATGCTGTACTACGTGTCTTAAAGTTTCATCAAAACAG	2100
Db	2041	ATTGGCCTCAGCGGTGCTTCATGATGCTGTACTACGTGTCTTAAAGTTTCATCAAAACAG	2100
OY	2101	AAACCACTCAAGACTGATTTACAGCCAGACGTCTGCGCGCTGTGGGGAGCCTGAC	2160
Db	2101	AAACCACTCAAGACTGATTTACAGCCAGACGTCTGCGCGCTGTGGGGAGCCTGAC	2160
OY	2161	AGACCTTTCAACTCAGCTCCCTCTCTCAGAGAGCCCTTCTGAGGACAGTGAAGGCAATGA	2220
Db	2161	AGACCTTTCAACTCAGCTCCCTCTCTCAGAGAGCCCTTCTGAGGACAGTGAAGGCAATGA	2220
OY	2221	CCCTACAGATGCTCAGCTACATCCGGCCAGGGGTCTGCACTGGCAGACAGCAAGCCACAG	2280
Db	2221	CCCTACAGATGCTCAGCTACATCCGGCCAGGGGTCTGCACTGGCAGACAGCAAGCCACAG	2280

Mon Jul 28 09:43:30 2003

us-09-989-981a-7.rnpb

Page 3

Db	2221	CCCTACAGATCTCAGCTACATCCGGGCCAGGGTGTCTCACTGGCACAGACCACCACAG	2280
QY	2281	GATGGCAGTAGAATAAAGACAGTCGAAAGGGATTTCTGCTACTGCGCAGAGACTGCGAT	2340
Db	2281	GATGGCAGTAGAATAAAGACAGTCGAAAGGGATTTCTGCTACTGCGCAGAGACTGCGAT	2340
QY	2341	GACTGGGAGAAAACCTGCATCGGTGGGCACCTACAGGTGCTAATTTATTTCCCTTTGA	2400
Db	2341	GACTGGGAGAAAACCTGCATCGGTGGGCACCTACAGGTGCTAATTTATTTCCCTTTGA	2400
QY	2401	TATGCATTTATATAGCACTCGATATATAGATGGGAGCAAACTAGAGATGATTTGGGTAG	2460
Db	2401	TATGCATTTATATAGCACTCGATATATAGATGGGAGCAAACTAGAGATGATTTGGGTAG	2460
QY	2461	CTAGACTGTGAGAAATTTGTGAACTGCGAGGAGAACATATACAGTAGCTAGCAGATTTG	2520
Db	2461	CTAGACTGTGAGAAATTTGTGAACTGCGAGGAGAACATATACAGTAGCTAGCAGATTTG	2520
QY	2521	GCTTCATCTTCCAGAGGGGCCCACTCCGTTGTAGCCACCATCATATACAGAAAGTGACC	2580
Db	2521	GCTTCATCTTCCAGAGGGGCCCACTCCGTTGTAGCCACCATCATATACAGAAAGTGACC	2580
QY	2581	TAAAGATGTACAGCAAGATGCATCCCTTCTTTTGTGTGGGGTCATAGGGCTCCAAAGC	2640
Db	2581	TAAAGATGTACAGCAAGATGCATCCCTTCTTTTGTGTGGGGTCATAGGGCTCCAAAGC	2640
QY	2641	CAACGTGACAACTTAAAAATGTATTTGAC	2669
Db	2641	CAACGTGACAACTTAAAAATGTATTTGAC	2669

RESULT 2

```

Sequence 3 Application US/09989981A
Publication No. US20030049730A1
GENERAL INFORMATION:
APPLICANT: Hobbs, Helen H.
APPLICANT: Shan, Bel
APPLICANT: Barnes, Robert
APPLICANT: Tian, Hui
APPLICANT: Tularix Inc.
APPLICANT: Board of Regents, The University of Texas System
TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
FILE REFERENCE: 018781-007320US
CURRENT APPLICATION NUMBER: US/09/989,981A
CURRENT FILING DATE: 2002-07-23
PRIOR APPLICATION NUMBER: US 60/252,235
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/253,645
PRIOR FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 2019
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(2019)
OTHER INFORMATION: mouse ABCG8 (mABCG8)
IS-09-989-981A-3

```

Query Match

Best Local Similarity 82.0%; Pred. No. 0;
Matches 1659; Conservative 0; Mismatches 360; Indels 3; Gaps 1.

QY 100 ATGCGCCGGGAGCGCGAGAGAGAGAGCGGCTGGCCGAAAGGGGCGACCTCCCGCAGGTATAC 159
 110 TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
Db 1 ATGCGCTGAGAAAACCAAAAGAAAGACCCAGCTGTGGATGGAGATGTACTTCCAGGATGCT 60
 110 TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
QY 160 TGGGGCCTCCAGGATAGATTCTTCTCTGAAAGATGACAACAGCCTGTACTTCCACCTAC 219
 170 TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
Db 61 TCGGGCCTCCAGGAGACGCTGTCTCTCTCGGAAAGTACAAACAGTCTGTACTTCCACCTAC 120
 130 TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT

QY	220	AGTGGCCAGCCCAACACCTCTGGAGGTGAGAGCCACATCTACCACTGAGTGCACCTGGCTCT	279
Db	121	AGTGTTCAGTCCAAACACTCTGGAGGTGACAGATCTCACTTACCAGTGGACATCGCTCT	180
QY	280	CAGGTCCCTTGTTGAGCAGCTGTGCTCAGTTCACAGATGCCCTGGACATCTCCACCTGC	339
Db	181	CAGGTGCTTGGTTGAGCAGCTGTGCTCAGTTCACAGATGCCCTGGAGGTCTCATAGCAGC	240
QY	340	CAGAAATCTTGTAGCTGGGCATCCAGAACTTAAGCTTCAAAAGTGAAGTGGGCAGATG	399
Db	241	CAAACTCTCTGTAGCTGGGCATCCGAAATCTAAGCTTCAAAAGTGAAGTGGAGCATG	300
QY	400	CTGGCCATCAATAGGAGAGCTCAGCTGTGTGGAGAGAGCTCTTGTCTAGATGTATCACTGGC	459
Db	301	CTGGCCATCAATAGGAGAGCTCAGGCTGCGGAGAAAGCTCACTCACTGACCTGTATCAAGGC	360
QY	460	CGAGGTCAAGGCGCGCAAGATCAAGTACAGGCCAGATCTGGATCAATAGGAGCGCAGCTGG	519
Db	361	AGAGGCCACGGTGGCAGATGAATACAGACAAATTTGGATTAATGGGCAACCCAGATAG	420
QY	520	CCTCAGCTGTGTAGGAAAGTGTGTGGCCCACTGTCGCACAGACAAACAGCTGTCTCCCAAC	579
Db	421	CCTCAGCTGTGTAGGAAAGTGTGGCTGTGCGATGTGTGGGAGCAATGACCAATCTGCTCCCAAC	480
QY	580	TTTGACTGTGCGAGAGACTTGGCCTTCAATTTGCCAGATGTGGGCTGCCAGAACTTCTCC	639
Db	481	CTGACCGGTCAAGAGAGACCTCGCTTCAATTTGCCAGATGTGGGCTGCCAGACCTTCTCC	540
QY	640	CAGGCCCAAGGCGTGAACAAAGGGTGGAGAGACGTGATCGGGAGACTCGGCTTAGCAGTGC	699
Db	541	CAGGCCCAAGGCGTGAACAAAGGGTGGAGAAACGTAAATCGCGAGCTCGGCTGTGGCAGTGC	600
QY	700	GCTGACACCCCGCTGGGGCAATGTACCTGTGGGGGGTGTGCGGGGGGTGAGCGCAGAGAA	759
Db	601	GCCAAACACAGTGGGCGAACACGATATAGCTGGGGGTGTCCGGGGGTGAGCGCGAGCA	660
QY	760	GTCAGCAATTGGGGGTGACAGCTCTGTGGAACCCAGGAATCTTATTTCTGCACAGAACCCACC	819
Db	661	GTGAGCAATTTGGGGGTGACAGCTCTGTGGAACCCAGGAATCTTATTTCTGATGTAACCCACT	720
QY	820	TCTTGGGCTTCACACACTTACACGCCACAACTGTGTGAAGACCTTGTCCAGCTGGCCAAA	879
Db	721	TCTTGGGCTTCGACACACTTACACGCCACAACTGTGTGAAGACCTTGTCCCGCTGGCCAAAG	780
QY	880	GGCAACCGGGGTGTCATCTCCCTCCACAGGCTGCTCGATCTTCAGGGCTGT	939
Db	781	GGCAACAGGCTGTGTCATCTCCCTCCACAGGCTGCTCGATCTTCAGGGCTGT	840
QY	940	GATCTGGTCTCTGATGACGCTGTGGACCCGCCATCTAATAGGGGCGGCCAGACATG	999
Db	841	GACCTGGTCTCTGATGACATCTGTGGACCCCTATCTAATAGGGGCGGCCAGACAAATG	900
QY	1000	GTCAGATATTTACAGCCATCGGCTAACCCCTGTCTCGCTACAGCAATCTGTGACTTTC	1059
Db	901	GTGAGATATTTACATCTCAATGAGCCACCCCTGTCTCGCTAATAGCAACCCCTCGGACCTTC	960
QY	1060	TATGTGACCTGCACACATGTAGACGGCGGACAGAGACAGAAATTTGGCCACAGGAG	1119
Db	961	TACGTGTGACTGTACACACATGTAGACGGCGGACAGAAAGGGAGTGTGGCCACCTGTGAG	1020
QY	1120	AAGGCTCAGTCACTGCGAGCCCTGTCTTCTAATAAAAGAGCTGTGACTTATGATGACTTCTGA	1179
Db	1021	AAGGCAACAAGTCTGTGAGCCCTGTCTTCTAATAAAAGTACAAAGCTTGTGATGACTTCTTG	1080
QY	1180	TGGAAAGCAGAGAGAAAGATCTTGTACGAGGACACTGTGTGGAAGACAGCGTGAACCCCA	1239
Db	1081	TGGAAAGCTGAGGCGAAAGAACTCAACAAAGCACCCACACATGTCAGCGCTGACCCCTACA	1140
QY	1240	CTTAGAACCAACATGCTCCCGAGTCTACAGAAATGCTGTGGGGGGGTGCAGCAGTATTACG	1299
Db	1141	CAGGACACTGACATG---TGGGACTGCTGTGTGAGCTGCCGGGAGTATAGACAGATTTTCC	1199

1100	1081	1240	1141
1080	1060	1040	1020
1000	980	960	940
920	900	880	860
840	820	800	780
760	740	720	700
680	660	640	620
600	580	560	540
520	500	480	460
440	420	400	380
360	340	320	300
280	260	240	220
200	180	160	140
120	100	80	60
40	20	0	

PRIOR APPLICATION NUMBER: 09/584,586
PRIOR FILING DATE: 2000-05-31
PRIOR APPLICATION NUMBER: PCT/US99/11825
PRIOR FILING DATE: 1999-05-27
PRIOR APPLICATION NUMBER: 60/086,988
PRIOR FILING DATE: 1998-05-28
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn version 3.0
SEQ ID NO 13
LENGTH: 2025
TYPE: DNA
ORGANISM: Mus musculus
US-09-866-866A-13

Query Match
Best Local Similarity 51.78; Score 122; DB 10; Length 2025;
Matches 278; Conservative 0; Mismatches 260; Indels 0; Gaps 0;

QY 538 TGTGTGGCCCGCCGCGCCAGCAGCAACACCTGCTCCCACTGAGCTGTGGAGAGACC 597
DB 374 TGTTCAGGTATATGTTGTTCAAGATGACGTGTGATGGGACCTGACAGTGAAGAAAC 433
QY 598 TTGGCCTTCAATTCAGATGCGGCTGCCAGAACCTTCGCCAGGCCGCGTGACAAA 657
DB 434 TTACAGTTCACAGCAGCTTCGACCTCCAAACAATATGAAGATCATGAATAAATGAA 493
QY 658 AGGGTGGAGAGCTGATCCGCGAGCTGCGGCTTAAAGCAGTGGCCTGACCCGCGTGGGC 717
DB 494 CGGATTAAACAACTATTAAGATAGGTAGCTGGAAAAAGTACCAATTTAAGGTGCGGA 553
QY 718 AACATGTAGTGGCGGGGTTGTGCGGGGTGAGCGAGAGAGTACAGCAATGGGGGCGAG 777
DB 554 ACACATTTATCCGTGCGCATCTGTGAGAGAAAGAAAGCAAGCANTAGAGATGAG 613
QY 778 CTCTCTGGAAACCCAGGAATCTTATTCGAGCAACCAACCTTGCGGCTCGACAGCTTC 837
DB 614 CTGATCACTGACCTTCACATCTCTCTGATGATGACCCAGCAGCTGTTTGGACTCAAG 673
QY 838 ACAGCCCAACACTGTTGAAGACTGTCCAGGCTGGCCAAAGGCAACGGGCTGTGCTC 897
DB 674 ACAGCAATGCTCTCTCTTGTCTCGAAAGGATCTTAACAGGGTGAACAAATCATC 733
QY 898 ATCTCCCTCCACAGCCTGCTGTGACATCTTCAGGCTGTTGATCTGGTCCCTCGATG 957
DB 734 TTCTCATTCATCAGCTGCTGCTGATTCATCTTAAGTTTGTGACAGCTCAGCTTACTG 793
QY 958 ACCTGTGCAACCCCACTTACTTAAAGGGGGGCCAGACATGATGATTTTCAAGCC 1017
DB 794 GCTTCGGGAAACTGTTTCCATGATGGCCAGACAGAAAGGCTTGAAGTACTTGTGATCA 853
QY 1018 ATCCGTACCCCTGCTGCTGCTACAGCAATCTCTGACTTCTATGTGAGCTGACCA 1075
DB 854 GCAGGTACCACTGTGAGCCCTTAACAACACCTGCGGATTTTCTTGATGTCAATCA 911

RESULT 8

US-09-866-866A-26
Sequence 26, Application US/09866866A
Patent No. US20020102244A1
GENERAL INFORMATION:
APPLICANT: Sorrentino, Brian
APPLICANT: Schuetz, John
TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
FILE REFERENCE: 1340-1-021CIP2
CURRENT APPLICATION NUMBER: US/09/866,866A
PRIOR FILING DATE: 2001-08-30
PRIOR APPLICATION NUMBER: 09/584,586
PRIOR FILING DATE: 2000-05-31
PRIOR APPLICATION NUMBER: PCT/US99/11825
PRIOR FILING DATE: 1999-05-27
PRIOR APPLICATION NUMBER: 60/086,988
PRIOR FILING DATE: 1998-05-28
NUMBER OF SEQ ID NOS: 27

SOFTWARE: PatentIn version 3.0
SEQ ID NO 26
LENGTH: 2247
TYPE: DNA
ORGANISM: Homo sapiens
US-09-866-866A-26

Query Match
Best Local Similarity 4.38; Score 115.4; DB 10; Length 2247;
Matches 269; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

QY 548 ACAGGCCCAACACACACCTGCTCCCACTGAGTGTGCGAGAGACCTTGGCTTCA 607
DB 561 ACCTGTACAGAGATGATGTTGTGATGGGCACTGAGCGGTGAGAAACCTTACAGTTCT 620
QY 608 TTGGCCAGATGCGGCTGCCAGAACCTTCTCCAGGCCCGACGCTGACAAAGGTGAGG 667
DB 621 CAGCAGCTCTTCCGCTTGCACAACTATGACATGATGATGATGATGATGATGATGATG 680
QY 668 ACCTGATCGGAGAGCTGCGGCTTAAAGCAGTGGCTGACACCCGCTGGGCAACATGACG 727
DB 681 GGGTCAATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 740
QY 728 TGGGGGGGTTGTGGGGGGTGGAGCGCAGAGAGTCAACATTTGGGTCAGCTCTCTGGA 787
DB 741 TCGGTGTGTGTGGAGGAGAAAGAAAGAGACTGATGATGATGATGATGATGATGATG 800
QY 788 ACCAGGAATCTTATTCGAGAACCCACCTGCGGCTGACAGCTTACAGCCACCA 847
DB 801 ATCTTCCATCTTCTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATG 860
QY 848 ACCTGTGAAGACCTTGTCCAGGCTGACCAAGCAACCGCTGCTGCTATCTCTCC 907
DB 861 CTGTCTTGTGTCTGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 920
QY 908 ACCAGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 967
DB 921 ATCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 980
QY 968 CCCCCATCTTAAAGGGGGGCCAGACATGATGATGATGATGATGATGATGATGATGATG 1027
DB 981 GATTAATGTTCCAGGCGCTGCTCAGAGAGCCCTTGGATGATGATGATGATGATGATG 1040
QY 1028 CCGTCTCTGCTACAGCAATCTGCTGACTTCTATGTGAGCTGA 1072
DB 1041 ACTGTAGGCTTATTAACCTGCAAGACTTCTTCTTGAGATCA 1085

RESULT 9

US-09-961-086-2
Sequence 2, Application US/09961086
Publication No. US2003003645A1
GENERAL INFORMATION:
APPLICANT: UNIVERSITY OF MARYLAND, BALTIMORE
APPLICANT: ROSS, Douglas D.
APPLICANT: DOYLE, L. Austin
APPLICANT: ABRUZZO, Lyne
TITLE OF INVENTION: BREAST CANCER RESISTANCE PROTEIN (BCRP) AND THE DNA
FILE REFERENCE: EP19376-019
CURRENT APPLICATION NUMBER: US/09/961,086
PRIOR FILING DATE: 2001-09-21
PRIOR APPLICATION NUMBER: US 60/073,763
PRIOR FILING DATE: 1998-02-05
PRIOR APPLICATION NUMBER: PCT/US99/02577
PRIOR FILING DATE: 1999-02-05
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 2418
TYPE: DNA
ORGANISM: Homo sapiens
US-09-961-086-2

[illegible]

Query Match	4.38	Score 115.4	DB 14	Length 2718
Best Local Similarity	51.23	Prod. No. 6.6e-24		
Matches 269	Conservative 0	Mismatches 256	Indels 0	Gaps 0
QY	548	ACGTGCGCCAGACACACACGACTGCTCCCAACTGACATGTGCAGAGACCTTGCGCTTCA	607	
Db	572	ACGTGTACCAAGATGATGTTGTATGATGGGCACTCTGACGGTGTGAGAAACTTACACACTTGT	631	
QY	608	TTGCCAGATGCGCGTCCAGAACCTTCTCCAGCGCCAGCGTGCACAAAGGGTGGAG	667	
Db	632	CACACAGCTTCTGGCTTGCACAAACATATGACGATCATGTAATAAAAAATGAAACGGATTATACA	691	

QY 668 ACGTATATCGGAGAGCTGGGGCTTAAAGCAGAGTCCGCGACACCCGCTGGGCAACATATGAC 727

Db 692 GGGTCATTCAAGAGATTAGGTCTGGATTAAATGAGCAGACTCCAAAGGTTTGGAACTCAAGTTA 751

QY 728 TCGGGGGGTTTCGGGGGGGTAGCGCAGAGAGAGTACAGCTATTTGGGTGACGCTCGTGGGA 787

Db 752 TCCGGGTGTGTCTGGAGAGAGAAAGAAAGAGACTAGTATAGATATGAGATGAGACGTTATCACTG 811

QY 788 ACCAAGGAATCTTATTTCGACAGCAACCCACTCTGGGCTGACAGCTTTCACAGCCACA 847

Db 812 ATCTTCACATCTTGTCTTGTGATGATGAGCCCTACACACTGCTTAGACTCAAGCAGACGAAATG 871

QY 848 ACTGTGTAAACCTGTGTCCAGGGCTGGGCGCAAGGCAACCGGCTGTGTCTATCTTCCCTCC 907

Db 872 CTGTCTCTTTTCTCTCTGAAAAGAGATGTAAAGCAGGAGCAACATCTATCTTCCATCTC 931

QY 908 ACCAGAGCTCGCTGACATCTTTCAGAGGCTGTTTGAATCTGTCTCTCGTATGACAGCTGGCA 967

Db 932 ATCAGCCCTGATATTCATCTTCAAGTATGTTTGAATAGGCTACCTTATTTGGGCTTCAGGAA 991

QY 968 CCCCACATCTATTAGGGGGCGGCCACGACATATGTCAGATATTTACAGAGCCATGGGCTACC.1027 1072

Db 992 GACTTATATGTTCCAGGGGCTGCTGCTCAGGAGGCGCTTGGGACTATTGTAATCAGACTGTATTC 1051

QY 1028 CCTTCCCTCGCTACAGCAATCTCTGACTCTTATGATGGAGACTCA 1072

Db 1052 ACTGTGAGGCTTATATATACCTCTCAGAGCTCTTCTTCTGGACATCA 1096

Db 909 TCAGGTGTGAGGCGCAAGCGCTGGCCATGCGCGCTGGAGCTGGTGAACAACCTCCAGTC 968
QY 799 CTATTTTCAGACCAACCACCTTGGGCTGAGACGTTCAACGCCACAACCTGTGAAG 858
Db 969 ATGTTCTTCATATGATGACCAACCGCTTGGACAGGCGCTCTGTTCCAGGTGCTCG 1028
QY 859 ACCCTGTCCAGGCTGGCCAAAGCAACCGGCTGGTGTCTATCTCCCTCCACCGCCTCG 918
Db 1029 CTATGTAAGAGGCTCGCTCAAGGGGGTGGCTTCATATTTGGACATCCACCGCCAGC 1088
QY 919 TCTGACATCTTACGCTGTTTGAATGTCCTCTGATGACGTCGTGACCCCATCTATC 978
Db 1089 GCCAACTCTTGAGCTGTTGACAGCTTTAGCTCTGATGATGACGACAAATGTGTAC 1148
QY 979 TTAGGGGGGGCCAGCAGCAATGTCAGATTTTCAAGCCATCGGCTATCCCTGCTCG 1038
Db 1149 CCGGGAAGATCTGCAATCTTGGCCATTTTGGAGATTTGGGATGTGAACCTCCCAAC 1208
QY 1039 TACAGCATCTGCTGCTGCTTCTATGAGACT--GACCAAGATTTACAGGGCCAGACA 1095
Db 1209 TACACACACCCAGCAGATTTTGTATGAGAGTTGCAATCCGGCAGTACGGTATCAGAAC 1268
QY 1096 GAGCAGGAATTTGGCCACCAAGGAGAGCTCACTGACAGCCCTGTTTCAAAAA 1155
Db 1269 AGTCGGCTGTGAGAGCGGTTGGGAGGCGATGTGACTGACACCAAGAGACCTTC 1328
QY 1156 GTGCGGACTTGTAGATGATCTTCTATGAAAGACAGACGAGATTTGACGAGACACC 1215
Db 1329 GGGGGTATGCGGAGGTGAACCTTTCTTTGGACCGGCGCTCTGACAGGATTAAGCAG 1388
QY 1216 TGTGTGAAGAGAGGTGACCCCACTAGACCACTGCTCCCGAGTCTTACGAAGATG 1275
Db 1389 ACMAAGATTTAAAGGGGTTGAAGAAAGACTCTCTCATGAAAGGCGCACAGCTTC 1448
QY 1276 CCTG---GGGCGGTGACGAGTTTACGAGCTGATCCGTGATTTTCAACGACTTC 1332
Db 1449 TGTGCGAGCTGCTACGAGAGTTTGTGATCTCTTCAAGAGACCTTCTCAGCATCAG 1508
QY 1333 CGAGACGTGCGCAACCTCTCTCATCTGAGGGGCGAGGCGCTGTGTATGTCATGACATC 1392
Db 1509 AAGGAGCTGCTCTGACACACCTGCGCATCCTGACATTTGGCATGGGCTCTCTCAT 1568
QY 1393 GGCCTTCCTATTTTGGCCATGAGGAGCATCCAGCTCTCTCATGATGATACAGCCCTTC 1452
Db 1568 GGCCTGCTGCTTACTTGGGAGTGGGAAACGAAGCCAAAGAGCTTTGAGCAACTCCGCTTC 1628
QY 1453 TTGTTCAAGTGGTCTCTCATCTCTTCAACGTCATTTGATGATCTATCTCAAAATG 1512
Db 1629 CTCTCTCTCTCATGCTGTTCTCTCAAGTTGCGGGCGCTCATGCTTACTGATCTT 1688
QY 1513 TACTAGAGAGGAGCAATGCTTACTATGAACTGGAAGAGGCGCTGTATACACACTGCTCA 1572
Db 1689 CCGCTGAGAGTGGAGGTCTTCTTTCGGAACACTGAACCTACTGCTAGAGCTGAAAGCC 1748
QY 1573 TATTTCTTTGCAAGATCCTGGGAGAGTCCGAGGAGCTGCTCATATCATCTATC 1632
Db 1749 TACTATCTGCGCAACATGAGCAACGCTGCTTTCAGATCATCTTCCAGTGGCTTAC 1808
QY 1633 GGGATGCCACCTACTGCTGCGCAACCTGAGGCGCTTCCAGGCTCTTCTGCTGAC 1692
Db 1809 TGCACATCTGCTACTGATGACGTCGAGCGCTGCGAGCGCGCTGCTGCTGCTGTT 1868
QY 1693 TTTCCGTGTGTGCTGCTGCTGCTCTTCTTCTGAGAGATTTAGGCGCTGCGCGCGGCGC 1752
Db 1869 GCGCGGCTGGGAGACCATCTCTCTGTTGGACAGTCCCTGCGCTCTGATGCGAGAC 1928
QY 1753 CTGCTCCCACTTTCACATGAGCTCTTCTTCAACAAAGCCCTTACAACTCTTAC 1812
Db 1929 GCGCTCAGCTCTGCTGAGTGGCGCTTCTGAGGCGCAAGTACAGCCATCCCGGTCTC 1988
QY 1813 CTGCGCGGCGCTTCTATGATAAACTTGAAGAGCTGTGAGACAGTCCCGCGGTGATTTCC 1872
Db 1989 CTGTTCTGCGGGTCTTCTGCTGAGCTTGCAGACAACTCCACAGTACTACAGTGGATGTC 2048

QY 1873 AAGTGTCTCTGCGGTGCTGTTTGAAGGCTCAT 1910
Db 2049 TACATCTCTATGTAGTATGCTTGAAGGCTCAT 2086

RESULT 14

US-09-954-531-591
Sequence 591, Application US/09954531
Patent No. US20020165180A1
GENERAL INFORMATION:
APPLICANT: Weaver, Zoe
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using
FILE OF INVENTION: Gene Sets
FILE REFERENCE: 689290-77
CURRENT FILING DATE: 2002-05-02
PRIOR APPLICATION NUMBER: US/09/954,531
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: US/60/234,009
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,034
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,509
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: US/60/234,567
PRIOR FILING DATE: 2000-09-22
NUMBER OF SEQ ID NOS: 1392
SOFTWARE: PatentIn version 3.0
SEQ ID NO 591
LENGTH: 2930
TYPE: DNA
ORGANISM: Homo sapiens
US-09-954-531-591

Query Match 3.9%; Score 105.2; DB 10; Length 2930;
Best Local Similarity 44.3%; Pred. No. 8.4e-21;
Matches 682; Conservative 0; Mismatches 823; Indels 33; Gaps 5;

QY 379 AAGTGAAGAGGGGCGACATGCTGCGCATCATAGGAGCTCAAGTTGGAGACCTTC 438
Db 337 AAGTTCAATGAGTGTGATGCTGTGTCATTAAGGCTCTTCCGGGCGGGAAGTCCAG 396
QY 439 TTGCTAGATGTGATCACTGCGGAGGTACAGCGGCGCAAGATCAAGTCAAGCCAGATGG 498
Db 397 CTGATGAACATCTGCTGCT-----GGATACAGAGGAGAGGGCATTAAGGGCGCTCTC 450
QY 499 ATCAATGGCGAGCCAGCTCGCTCAGCTGTGAGAGAGTGTGAGCCAGCTGCGCAG 558
Db 451 ATCAAGGCGCTGCGCGGAGCTGCGCTGCTTCCGGAAGTGTCTGCTACATCATGAG 510
QY 559 CACAAAGCTGCTGCGCCCACTGATGCTGTGAGAGAGCTTGGCTTCAATGCGCAGATG 618
Db 511 GATGACATGCTGCTGCGCATCTCTGAGAGGCGCATATGATGTGCGCATATG 570
QY 619 CCGCTGCCAGAACTTCTCCAGGCGCGGTGACAAAGAGGTGAGAGCTGATGCGC 678
Db 571 AAGCTTCAGGAAGG-----ATGAAGCAGAGAGGAAATGTCTAAGAGATATGACA 624
QY 679 GAGCTGCGGCTTAGCACTGCTGACACCCGCGTGGCAACATGATACGTGCGGGGTTG 738
Db 625 GCGCTGCGGCTGCTGCTTGTGCGCAACGCGGAGACG-----GGAGCTG 669
QY 739 TCGGGGGTGTGAGCGAGAGAGTACAGATTTGGGGTGCAGCTCTGTGAGAACAGATC 798
Db 670 TCAGTGTGTAGCGGCAAGCGCTGCGCTGCGCTGAGGCTGTGTAACAACTTCCATC 729
QY 799 CTATTTCTGAGGAACCACTTGTGGCTGACAGCTTCAAGGCCAACAACTGTGTGAAG 858
Db 730 ATGTTCTTGATGAGCCACACAGGCGCTGAGAGCGCTCTCTTCCAGTGTCTCG 789
QY 859 ACCTTTCCAGGTGCGCAAGGCAACCGGCTGTGCTCATCTCTTCCACCAAGCTTCC 918

Db 790 CTGATGAAGGCGCTGCTCAAGGGGCTGCTCCATCATATTTGCAACCATCCACGCCAGC 849
 Oy 919 TGTGACATCTTCAGAGCTTTTGTATGATGTCCTCTGATGACGTCGACCCCATCTAC 978
 Db 850 GCCAAACTCTTCGAGCTTTCGACACGCTTTAGCTCTGATGACAGCAATGCTGTAC 909
 Oy 979 TTAGGGGCGGCCAGCAGCATGCTCAGTATTTACAGCCATCGGCTACCCCTGCTCCG 1038
 Db 910 CGGGGAAAGCTTGCATCTTGTGCTATTTGAGGATTTGGGTGTGAACTGCCAAC 969
 Oy 1039 TACAGCAATCTGCTGCTCTATGTCGACCT---GACCAGCATTTGACAGCCGACGAGA 1095
 Db 970 TACCAACACCCAGCAGATTTTGTATGATGAGGCTGATCGGCGAGTACGGTGTATCAAC 1029
 Oy 1096 GAGCAGCAATTTGCGCACAGGAGAGAAAGCTCATGCTCCAGCCCTGTTTCAAGAAA 1155
 Db 1030 AGTCGGCTGTGAGAGGGGTTGCGAGGGCATGTGTACACACCAACAGAGAGACCTC 1089
 Oy 1156 GTGCGTGTACTTATGATGATCTTTATGAGAAAGAGAGAGAGATCTTGAAGAGAC 1215
 Db 1090 GGGGTGTATGCCAGGTGAACCTTTCTTTGGCACCGCCCTCTGTAAGAGTAAAGCAG 1149
 Oy 1216 TGTGTGAAGAGCGGTGACCCGACTAGACACCACTGCTCCGAGTCTTACGAGATG 1275
 Db 1150 ACAAAGATTAAGGGGTTGAGAAAGAGTCTCTGTCATGAGAGGCTGCCACAGCTTC 1209
 Oy 1276 CCTG---GGCGGTGAGCAGATTACGACGCTGATCGTGTGATTTCCAGAGCTTC 1332
 Db 1210 TGTGACAGCTGCTCAGGAGCTTCTGATCTTCAAGAGACCTTCTCAGCATCAG 1269
 Oy 1333 CGAGACCTTGCACCTCTCATATCAGGGGCGAGGCTGTGTATGTCAATGACCATC 1392
 Db 1270 AGGAGCTGCTCTGACACACCTGCGCATCTGCAATGCGCATTTGGGCTCTGCTCAT 1329
 Oy 1393 GGGTCTCTATTTTGGGATGGGAGCATTCAGCTCTCTCATGATACAGCCGCTTC 1452
 Db 1330 GGGCTCTGCTGCTGAGATCGGGAACGAAACAGAAAGGTCTTGGAGCACTCCGCTTC 1389
 Oy 1453 TTGTCATATGCTGCTCTCATCTCTTCAACGATCATCTGATGATTCATCTCAATG 1512
 Db 1390 CTCTTCTCTCATGCTGCTTCTCATGTTGCGGCGCTTCAATGCTTCTGATCATTT 1449
 Oy 1513 TACTGAGAGAGGCAATGCTTACTATGAACTGGAAGAGGCGCTGTACACCACTGCTCA 1572
 Db 1450 CCCCTGAGATGGAGCTTTCTTCTGGAACACCTGACATGCTGATGACAGCCGAGAG 1509
 Oy 1573 TATTTCTTCCAAAGTCTCGGGAGCTTCGAGACCTGTGCTTACATCATCTATC 1632
 Db 1510 TACTACTGCGCCAAAGCAATGAGAGAGCTTCCATGATCATGTTCCAGAGGCTAC 1569
 Oy 1633 GGGATCCCACTTACGCTGCGGCAACGTCGAGGCGAGGCTCCAGCCCTTCTGCTGCA 1692
 Db 1570 TGCAGATGCTGATGAGATGAGAGTGCACCGCTCGAGCGCTTGTGCTGCTGTT 1629
 Oy 1693 TTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1752
 Db 1630 GCGGCGCTGGGACATGACCTCCGTCGTCGACATGCTCCCTGCGCTGCTGATCGAG 1689
 Oy 1753 CTGCTCCCACTTGCACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1812
 Db 1690 GCTCTCAGCTCCCTGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1749
 Oy 1813 CTGCGCGGGGCTTCTGATTAACCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1872
 Db 1750 CTGCTTCTGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1809
 Oy 1873 AAAGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1910
 Db 1810 TACATCTCTATGCTAGGTATGGGTTGAGAGGGGCTAT 1847

RESULT 15
 US-10-171-581-276

Sequence 276, Application US/10171581
 Publication No. US20030104426A1
 GENERAL INFORMATION:
 APPLICANT: Dai, Hongyue
 APPLICANT: Linsley, Peter
 APPLICANT: Mao Mao
 TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia
 FILE REFERENCE: 9301-157-999
 CURRENT APPLICATION NUMBER: US/10/171,581
 PRIOR FILING DATE: 2002-06-14
 PRIOR APPLICATION NUMBER: 60/298,914
 NUMBER OF SEQ ID NOS: 366
 SEQ ID NO 276
 LENGTH: 2930
 TYPE: DNA
 ORGANISM: Homo sapiens
 PUBLICATION INFORMATION:
 DATABASE ACCESSION NUMBER: X91249
 DATABASE ENTRY DATE: 2001-06-18
 US-10-171-581-276

Query Match 3.9%; Score 105.2; DB 14; Length 2930;
 Best Local Similarity 44.3%; Pred. No. 8.4e-21;
 Matches 682; Conservative 0; Mismatches 823; Indels 33; Gaps 5;

Oy 379 AAAGTGAAGGTGGGAGATGCTGGCCATCATAGGAGCTCAGGTTGTGGAGAGCTTC 438
 Db 337 AAGTTCAATGTTGATGATGTTGAGGCTATTTGCTTCCGCGGCGGGAAGTCCACG 396
 Oy 439 TTGCTGATGATGATCTGCGCGGAGGTCACGGGCGGAGATCAAGTCAGGCGCATCTG 498
 Db 397 CTGATCAATCTCTGCT-----GGATACAGGAGAGCGGCAATGAGGGGCGCTCTC 450
 Oy 499 ATCAATGAGGAGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 558
 Db 451 ATCAAGGCGGCTGCGGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 510
 Oy 559 CACACACCTGCTCCCACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 618
 Db 511 GATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 570
 Oy 619 CGGCTGCCAGAACTTCTCCAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 678
 Db 571 AAGCTTCAGAGAGAG-----ATGAGGAGAGAGGAGGAGGAGGAGGAGGAGGAGG 624
 Oy 679 GAGCTGCGGCTTACGAGTGGCTGACACCGCGCTGGGCAACATGATGCTGCGGGGTTG 738
 Db 625 GCGCTGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 669
 Oy 739 TCGGGGGTGAAGGAGGAGTACAGATGATGATGATGATGATGATGATGATGATGATG 798
 Db 670 TTAGGTGTGACAGGAGGAGGCGCTGCGCTGAGTGGAGTGGAGTGGAGTGGAGTGG 729
 Oy 799 CTATTTCTGAGAGAACCCACCTCTGCGCTGAGAGGATTCAGAGCCCAACCTGCTGAG 858
 Db 730 ATGTTCTTGAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 789
 Oy 859 ACCCTGTCAGGCTGCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 918
 Db 790 CTGATGAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 849
 Oy 919 TGTGATCTTGAAGCTGTTTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 978
 Db 850 GCCAAACTCTGAGCTGTTGACCAAGCTTACGTCCTGAGTCAAGGACAAATGCTGTAC 909
 Oy 979 TTAGGGGCGGCGCACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1038
 Db 910 CGGGGAAAGTGTCAATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 969
 Oy 1039 TACAGCAATCTGCTGCTTATGTCGACCT---GACCAGATTCAGAGGCGGAGCAGCA 1095
 Db 970 TACCAACACCCAGCAGATTTTGTCTATGAGGTTGATCTGCGGCGAGTACGGTATCA 1029

Search completed: July 27, 2003, 08:56:18
Job time : 583.074 secs

Query Match	1.9%;	Score 51.4;	DB 4;	Length 1155,
Best Local Similarity	46.0%;	Pred. No. 0.0029;		

OY	468	CGGGGGAATATCAAGTCAGGCCAGATCTGGATTAATGGGCAGCCCACTTGCCTCAACT	527
Dd	1569	GCGCGGACACACACACATCAAGCTGGTTCTTGCGCTGTGGCCCCCACAAGAGCGG	1628
OY	528	GGTGAGAAAGTGTGTGSCCACACTGCGCCACACACCAGCTGTGTCCCAATTGACTGT	587
Dd	1629	CGTCCGGGTCTCGGCACCATGTGGAGAAGCTCTGAAGCGCGCGCCACAGCTCGGCTACT	1688
OY	588	GCGAGAACCTTGGCCCTTCAATTGCCAGATGCGGCTGCCCGAACACTTCTCCAGGGCCA	647
Dd	1689	GCGGGAAGAGTCACTTCTACCGCAGCTCAGTGGGGCGGGAAACCTTGAGGTCACTTTCG	1748

RESULT 13
US-08-592-874-1/c
; Sequence 1, Application US/08592874
; Patent No. 5854034

```

: ADDRESS: JULES E. GOLDBERG
: STREET: 261 MADISON AVENUE
: CITY: NEW YORK
: STATE: NY
: COUNTRY: USA
: ZIP: 10016-2391
:

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.300
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,874
FILING DATE:

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PG 08/277 440

```

FILED DATE: 24-JAN-1995
ATTORNEY/AGENT INFORMATION:

NAME: GOLDBERG, JULES E.
REGISTRATION NUMBER: 24,408
TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-986-4090
TELEFAX: 212-818-9479
INFORMATION FOR PROSPECTORS

```

; INFORMATION FOR SEQ ID NO: 1:
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 28804 base pairs

```

TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)
FRAGMENT TYPE: N-terminal

Query Match	1.9%	Score 51.4;	DB 2;	Length 28804;
Best Local Similarity	47.9%;	Pred. No. 0.0094;		
Matches 148; Conservative	0;	Mismatches 161;	Indels 0;	Gaps 0

RESULT 14
US-09-096-942-2/c
; Sequence 2, Application US/09096942
; Patent No. 6037025

PATENT NO.: 0027923
 GENERAL INFORMATION:
 APPLICANT: Pollock, Thomas J
 APPLICANT: Mikolajczak, Marcia
 APPLICANT: Yamazaki, Motonide
 APPLICANT: Thorne, Linda
 APPLICANT: Amentrout, Richard W
 TITLE OF INVENTION: Production of Xanthan Gum by *Sphingomonas Bacteri*
 TITLE OF INVENTION: Carrying Genes from *Xanthomonas Campestris*
 FILE REFERENCE: seq list for appl filed from pro. appl
 CURRENT APPLICATION NUMBER: US/09/096,942

; CURRENT FILING DATE: 1998-06-12
 ; EARLIER APPLICATION NUMBER: 60/049,428
 ; EARLIER FILING DATE: 1997-06-12
 ;

```

; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2

```

```

; LENGTH: 28804
; TYPE: DNA
; ORGANISM: Schistosoma

```

US-09-096-942-2

Query Match	1.98;	Score 51.4;	DB 3;	Length 28804;
Best Local Similarity	47.98;	Pred. No. 0.0094;		
Matches 148; Conservative	0;	Mismatches 161;	Indels 0;	

606 CATTGCCAGATGGGCTGCCAGAACCTTCTCCAGGCCCGCGGTGACAAAGGTTGA 665

D_b 19351 CATTGGGCTGTCCAACCCGGGATGCGCGTTCGAGCATGTCTGCTGGCGGGCGGACGCTGGC 19292

666 GGAAGCTGATCGCGGAGCTGGGGCTTAGGCAAGTGGCGTGGACACCCGGCTGGGCAATGTATA 725

726 CGTGGCGGGGTTGTCGGGGGCTGACGCGACAGATCAGCATTTGGGCTGCACCTCCTGTG 785

Db 19231 CGGCGTCAACCTGTCTGGGGCGGCCAGCGCCAGCGGCTTCGCTATCGCCCGCGGCGCTGGTGG 19172

QY 786 GAACCCAGGAATCTTATTTCTCGACGAACCCACCTTCGGGCTCGACAGCTTCACAGCCCA 845

846 CAACCTGGTGAAGACTTTGTCAGAGCTGGCTAAAGCACC66GTGTCATCTCCCT 905
 1371 TCGAATCCGCGCATCCATGGIGTILGAGAGAGGCGCACCTCCGCGCTGAGATGCCGAGAGCGGAGA 1911Z

Db 19111 GCTGATCCAGAACATCTTCGGCGCATCTTCGGCGGCGCGACGCTGTGATCATCGCCCA 19052

Oy	906	CCACCAGCC	914
Db	19051	CCGCGTGTC	19043

RESULT 15

US-09-096-867-2/c
; Sequence 2, Application US/09096867
; Patent No. 6020817

GENERAL INFORMATION:

APPLICANT: Pollock, Thomas J

APPLICANT: Mikolajczak, Marcia

APPLICANT: Thorne, Linda

APPLICANT: Armentrout, R

TITLE OF INVENTION: Production of

1. TITLE OF INVENTION: Carrying Genes from Xanthomonas Campestris
2. PRIOR ART REFERENCES: 1. U.S. Pat. 4,014,500 (1977) 2. U.S. Pat. 4,014,501 (1977) 3. U.S. Pat. 4,014,502 (1977) 4. U.S. Pat. 4,014,503 (1977) 5. U.S. Pat. 4,014,504 (1977) 6. U.S. Pat. 4,014,505 (1977) 7. U.S. Pat. 4,014,506 (1977) 8. U.S. Pat. 4,014,507 (1977) 9. U.S. Pat. 4,014,508 (1977) 10. U.S. Pat. 4,014,509 (1977) 11. U.S. Pat. 4,014,510 (1977) 12. U.S. Pat. 4,014,511 (1977) 13. U.S. Pat. 4,014,512 (1977) 14. U.S. Pat. 4,014,513 (1977) 15. U.S. Pat. 4,014,514 (1977) 16. U.S. Pat. 4,014,515 (1977) 17. U.S. Pat. 4,014,516 (1977) 18. U.S. Pat. 4,014,517 (1977) 19. U.S. Pat. 4,014,518 (1977) 20. U.S. Pat. 4,014,519 (1977) 21. U.S. Pat. 4,014,520 (1977) 22. U.S. Pat. 4,014,521 (1977) 23. U.S. Pat. 4,014,522 (1977) 24. U.S. Pat. 4,014,523 (1977) 25. U.S. Pat. 4,014,524 (1977) 26. U.S. Pat. 4,014,525 (1977) 27. U.S. Pat. 4,014,526 (1977) 28. U.S. Pat. 4,014,527 (1977) 29. U.S. Pat. 4,014,528 (1977) 30. U.S. Pat. 4,014,529 (1977) 31. U.S. Pat. 4,014,530 (1977) 32. U.S. Pat. 4,014,531 (1977) 33. U.S. Pat. 4,014,532 (1977) 34. U.S. Pat. 4,014,533 (1977) 35. U.S. Pat. 4,014,534 (1977) 36. U.S. Pat. 4,014,535 (1977) 37. U.S. Pat. 4,014,536 (1977) 38. U.S. Pat. 4,014,537 (1977) 39. U.S. Pat. 4,014,538 (1977) 40. U.S. Pat. 4,014,539 (1977) 41. U.S. Pat. 4,014,540 (1977) 42. U.S. Pat. 4,014,541 (1977) 43. U.S. Pat. 4,014,542 (1977) 44. U.S. Pat. 4,014,543 (1977) 45. U.S. Pat. 4,014,544 (1977) 46. U.S. Pat. 4,014,545 (1977) 47. U.S. Pat. 4,014,546 (1977) 48. U.S. Pat. 4,014,547 (1977) 49. U.S. Pat. 4,014,548 (1977) 50. U.S. Pat. 4,014,549 (1977) 51. U.S. Pat. 4,014,550 (1977) 52. U.S. Pat. 4,014,551 (1977) 53. U.S. Pat. 4,014,552 (1977) 54. U.S. Pat. 4,014,553 (1977) 55. U.S. Pat. 4,014,554 (1977) 56. U.S. Pat. 4,014,555 (1977) 57. U.S. Pat. 4,014,556 (1977) 58. U.S. Pat. 4,014,557 (1977) 59. U.S. Pat. 4,014,558 (1977) 60. U.S. Pat. 4,014,559 (1977) 61. U.S. Pat. 4,014,560 (1977) 62. U.S. Pat. 4,014,561 (1977) 63. U.S. Pat. 4,014,562 (1977) 64. U.S. Pat. 4,014,563 (1977) 65. U.S. Pat. 4,014,564 (1977) 66. U.S. Pat. 4,014,565 (1977) 67. U.S. Pat. 4,014,566 (1977) 68. U.S. Pat. 4,014,567 (1977) 69. U.S. Pat. 4,014,568 (1977) 70. U.S. Pat. 4,014,569 (1977) 71. U.S. Pat. 4,014,570 (1977) 72. U.S. Pat. 4,014,571 (1977) 73. U.S. Pat. 4,014,572 (1977) 74. U.S. Pat. 4,014,573 (1977) 75. U.S. Pat. 4,014,574 (1977) 76. U.S. Pat. 4,014,575 (1977) 77. U.S. Pat. 4,014,576 (1977) 78. U.S. Pat. 4,014,577 (1977) 79. U.S. Pat. 4,014,578 (1977) 80. U.S. Pat. 4,014,579 (1977) 81. U.S. Pat. 4,014,580 (1977) 82. U.S. Pat. 4,014,581 (1977) 83. U.S. Pat. 4,014,582 (1977) 84. U.S. Pat. 4,014,583 (1977) 85. U.S. Pat. 4,014,584 (1977) 86. U.S. Pat. 4,014,585 (1977) 87. U.S. Pat. 4,014,586 (1977) 88. U.S. Pat. 4,014,587 (1977) 89. U.S. Pat. 4,014,588 (1977) 90. U.S. Pat. 4,014,589 (1977) 91. U.S. Pat. 4,014,590 (1977) 92. U.S. Pat. 4,014,591 (1977) 93. U.S. Pat. 4,014,592 (1977) 94. U.S. Pat. 4,014,593 (1977) 95. U.S. Pat. 4,014,594 (1977) 96. U.S. Pat. 4,014,595 (1977) 97. U.S. Pat. 4,014,596 (1977) 98. U.S. Pat. 4,014,597 (1977) 99. U.S. Pat. 4,014,598 (1977) 100. U.S. Pat. 4,014,599 (1977) 101. U.S. Pat. 4,014,600 (1977) 102. U.S. Pat. 4,014,601 (1977) 103. U.S. Pat. 4,014,602 (1977) 104. U.S. Pat. 4,014,603 (1977) 105. U.S. Pat. 4,014,604 (1977) 106. U.S. Pat. 4,014,605 (1977) 107. U.S. Pat. 4,014,606 (1977) 108. U.S. Pat. 4,014,607 (1977) 109. U.S. Pat. 4,014,608 (1977) 110. U.S. Pat. 4,014,609 (1977) 111. U.S. Pat. 4,014,610 (1977) 112. U.S. Pat. 4,014,611 (1977) 113. U.S. Pat. 4,014,612 (1977) 114. U.S. Pat. 4,014,613 (1977) 115. U.S. Pat. 4,014,614 (1977) 116. U.S. Pat. 4,014,615 (1977) 117. U.S. Pat. 4,014,616 (1977) 118. U.S. Pat. 4,014,617 (1977) 119. U.S. Pat. 4,014,618 (1977) 120. U.S. Pat. 4,014,619 (1977) 121. U.S. Pat. 4,014,620 (1977) 122. U.S. Pat. 4,014,621 (1977) 123. U.S. Pat. 4,014,622 (1977) 124. U.S. Pat. 4,014,623 (1977) 125. U.S. Pat. 4,014,624 (1977) 126. U.S. Pat. 4,014,625 (1977) 127. U.S. Pat. 4,014,626 (1977) 128. U.S. Pat. 4,014,627 (1977) 129. U.S. Pat. 4,014,628 (1977) 130. U.S. Pat. 4,014,629 (1977) 131. U.S. Pat. 4,014,630 (1977) 132. U.S. Pat. 4,014,631 (1977) 133. U.S. Pat. 4,014,632 (1977) 134. U.S. Pat. 4,014,633 (1977) 135. U.S. Pat. 4,014,634 (1977) 136. U.S. Pat. 4,014,635 (1977) 137. U.S. Pat. 4,014,636 (1977) 138. U.S. Pat. 4,014,637 (1977) 139. U.S. Pat. 4,014,638 (1977) 140. U.S. Pat. 4,014,639 (1977) 141. U.S. Pat. 4,014,640 (1977) 142. U.S. Pat. 4,014,641 (1977) 143. U.S. Pat. 4,014,642 (1977) 144. U.S. Pat. 4,014,643 (1977) 145. U.S. Pat. 4,014,644 (1977) 146. U.S. Pat. 4,014,645 (1977) 147. U.S. Pat. 4,014,646 (1977) 148. U.S. Pat. 4,014,647 (1977) 149. U.S. Pat. 4,014,648 (1977) 150. U.S. Pat. 4,014,649 (1977) 151. U.S. Pat. 4,014,650 (1977) 152. U.S. Pat. 4,014,651 (1977) 153. U.S. Pat. 4,014,652 (1977) 154. U.S. Pat. 4,014,653 (1977) 155. U.S. Pat. 4,014,654 (1977) 156. U.S. Pat. 4,014,655 (1977) 157. U.S. Pat. 4,014,656 (1977) 158. U.S. Pat. 4,014,657 (1977) 159. U.S. Pat. 4,014,658 (1977) 160. U.S. Pat. 4,014,659 (1977) 161. U.S. Pat. 4,

FILE REFERENCE: seq list for appl tiled from pro. appl
CURRENT APPLICATION NUMBER: NS/09/006 857

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

EARLIER APPLICATION NUMBER: 60/049,428

EARLIER FILING DATE: 1997-06-12

```

; NUMBER OF SEQ ID NOS: 2
SOFTWARE: DnaSP v1.6.0

```

SOFTWARE: PATENTLIN Ver. 2.0
: SEO ID NO 2

LENGTH: 28804

TYPE: DNA

ORGANISM:

US-09-096-867-2

Query Match

Query Match	1.9%	Score 51.4	DB 3	Length 28804
Best Local Similarity	47.9%	Pred. No. 0.0094		
Matches 148	Conservative	0	Mismatches 161	Indels 0
				Gaps 0

666 GGACGTGATCGCGAAGCTGCGGCTTAAGCAGTGGGCTGACACCCCGGTGGGCAACATGTA 725

Dd 19231 CGGGCTAACCTGTGGGGGGGCCAGCGGCTTCGTCGCCCGCGCTGTCGG 19172

846 CAACCTGTCAGACCTTGTCCAGGCTGGCCAAAGGCAACCGGCTGTCTCATCTCCT 905

19031 CCGC167C 19043

Search completed: July 27, 2003, 12:45:54
Job time : 182.378 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 09:04:42 ; Search time 15 Seconds

(without alignments)

1800.073 Million cell updates/sec

Title: US-09-989-981a-6

Perfect score: 3326

Sequence: 1 MCDLSSLTFCGSMGLQVNRG.....PALVILGIIVKIRDLISR 651

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3326	100.0	651	1	ABG5_HUMAN
2	2738.5	82.3	652	1	ABG5_MOUSE
3	2727.5	82.0	652	1	ABG5_RAT
4	698	21.0	673	1	ABG8_MOUSE
5	697	21.0	673	1	ABG8_HUMAN
6	691	20.8	672	1	ABG8_RAT
7	676.5	20.3	655	1	ABG2_HUMAN
8	627	18.9	1294	1	YOH5_YEAST
9	623	18.7	677	1	WHIT_LOCCU
10	621	18.7	1049	1	ADP1_YEAST
11	607.5	18.3	695	1	WHIT_ANOGA
12	602.5	18.1	687	1	WHIT_DROME
13	596.5	17.9	678	1	ABG1_HUMAN
14	591	17.8	679	1	WHIT_CERCA
15	578.5	17.4	646	1	ABG4_HUMAN
16	573	17.2	709	1	WHIT_ANOAL
17	569.5	17.1	666	1	ABG1_MOUSE
18	561.5	16.9	598	1	YPC3_CAEEL
19	547	16.4	666	1	SCRT_DROME
20	517.5	15.6	610	1	YOSC_CAEEL
21	454.5	13.7	675	1	BROW_DROME
22	435	13.1	668	1	BROW_DROVI
23	427	12.8	1499	1	CDR2_CANAL
24	406.5	12.2	1333	1	YNS9_YEAST
25	403.5	12.1	1364	1	PDR8_YEAST
26	399.5	12.0	1530	1	BFR1_SCHPO
27	398	12.0	1501	1	SNQ2_YEAST
28	392.5	11.8	1501	1	CDR3_CANAL
29	391	11.8	1529	1	PDRF_YEAST
30	383.5	11.5	650	1	ABG3_MOUSE
31	382	11.5	1501	1	CDR1_CANAL
32	382	11.5	1511	1	PDR8_YEAST
33	374	11.2	1511	1	PDR5_YEAST

ALIGNMENTS

RESULT 1	ABG5_HUMAN	STANDARD:	PRT:	651 AA.
AC	Q9H222;			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	ATP-binding cassette, sub-family G, member 5 (Sterolin-1).			
GN	ABG5.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID:9606;			
RP	[1]			
RP	SEQUENCE FROM N.A., AND VARIANT GLD-604.			
RC	TISSUE=Liver;			
EX	MEDLINE-20553648; PubMed-11099417;			
RA	Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,			
RA	Kwiterovich P., Shan B., Barnes R., Hobbs H.H.;			
RT	"Accumulation of dietary cholesterol in sitosterolemia caused by			
RT	mutations in adjacent ABC transporters.";			
RL	Science 290:1771-1775(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A., VARIANTS SITOSTEROLEMIA H-389; H-419 AND P-419,			
RP	AND VARIANT E-604.			
RC	TISSUE=Liver;			
RX	MEDLINE-20578753; PubMed-11138003;			
RA	Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,			
RA	Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,			
RA	Dean M., Patel S.B.;			
RT	"Identification of a gene, ABCG5, important in the regulation of			
RT	dietary cholesterol absorption.";			
RL	Nat. Genet. 27:79-83(2001).			
RN	[3]			
RP	REVIEW.			
RP	MEDLINE-21474438; PubMed-11590207;			
RA	Schmitz G., Langmann T., Heimerl S.;			
RT	"Role of ABCG1 and other ABCG family members in lipid metabolism.";			
RL	J. Lipid Res. 42:1513-1520(2001).			
RN	[4]			
RP	VARIANTS SITOSTEROLEMIA Q-146; H-389; P-419; H-419 AND S-550, AND			
RP	VARIANT E-604.			
RX	MEDLINE-21344600; PubMed-11525359;			
RA	Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,			
RA	Ose U., Stenhoef A.F.H., Miettinen T., Bjorkhem I., Bruckert E.,			
RA	Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,			
RA	Patel S.B.;			
RT	"Two genes that map to the STSL locus cause sitosterolemia: genomic			
RT	structure and spectrum of mutations involving sterolin-1 and			
RT	sterlin-2, encoded by ABCG5 and ABCG8, respectively.";			
RL	Am. J. Hum. Genet. 69:278-290(2001).			
CC	-1- FUNCTION: Transporter that appears to play an indispensable role			
CC	of the enterocytes and in the selective sterol excretion by the			
CC	liver into bile.			
CC	-1- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to			

34	367	11.0	1410	1	PDR8_YEAST	P40550 saccharomyc
35	365.5	11.0	1490	1	CDR4_CANAL	O74676 candida alb
36	263.5	7.9	345	1	ABG5_HAEIN	P44785 haemophilus
37	263	7.9	241	1	XHBG_HAEIN	P45073 haemophilus
38	257	7.7	365	1	CYSA_ECOLI	P46676 escherichia
39	254	7.6	365	1	CYSA_SATY	P40860 salmonella
40	251	7.5	231	1	CYDC_ECOLI	P23886 escherichia
41	250.5	7.5	344	1	YTRF_BACSU	O34392 bacillus su
42	249.5	7.5	344	1	CYSA_SYMP7	P14788 synechococc
43	247.5	7.4	356	1	HITC_HAEIN	P44513 haemophilus
44	241.5	7.3	355	1	CYSA_SYMY3	P74548 synechocyst
45	240.5	7.2	308	1	NOSF_PSEST	P19844 pseudomonas

CC	ABCG8 along a pathway regulating dietary-sterol absorption and excretion.
CC	-1 SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC	-1 TISSUE SPECIFICITY: Strongly expressed in the liver, lower levels in the small intestine and colon.
CC	-1 DISEAS: Defects in ABCG5 are a cause of sitosterolemia, also known as phytosterolemia or shellfish sterolemia, a rare autosomal recessive disorder characterized by increased intestinal absorption of all sterols including cholesterol, plant and shellfish sterols, and decreased biliary excretion of dietary sterols into bile. Sitosterolemia patients have hypercholesterolemia, very high levels of plant sterols in the plasma, and frequently develop tendon and tuberosus xanthomas, accelerated atherosclerosis and premature coronary artery disease.
CC	-1 SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE) SUPERFAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/or_send_an_email_to_license@sib-sib.ch).
CC	-----
DR	EMBL; AF320293; AAC40003.1; -
DR	EMBL; AF312715; AAC53099.1; -
DR	GeneW; HGNC:13886; ABCG5.
DR	MM; 605459; -
DR	MM; 210250; -
DR	InterPro; IPR003593; AAA_ATPase.
DR	InterPro; IPR003439; ABC_transportr.
DR	Pfam; PF00005; ABC_tran; 1.
DR	ProDom; PD000006; ABC_transportr; 1.
DR	SMART; SM00382; AAA; 1.
DR	PROSITE; PS00211; ABC_TRANSPORTER; FALSE_NEG.
KM	Arp-binding; Glycoprotein; Transmembrane; Transport; Polymorphism; Disease mutation.
KW	DOMAIN
FT	DOMAIN 1 383
FT	TRANSMEM 384 404
FT	DOMAIN 405 421
FT	TRANSMEM 422 442
FT	DOMAIN 443 462
FT	TRANSMEM 463 483
FT	DOMAIN 484 503
FT	TRANSMEM 504 524
FT	DOMAIN 525 528
FT	TRANSMEM 529 549
FT	DOMAIN 550 623
FT	TRANSMEM 624 644
FT	DOMAIN 645 651
FT	NP_BIND 86 93
FT	CARBOHYD 584 584
FT	CARBOHYD 591 591
FT	VARIANT 146 146
FT	FT
FT	VARIANT 389 389
FT	FT
FT	VARIANT 419 419
FT	FT
FT	VARIANT 419 419
FT	FT
FT	VARIANT 550 550
FT	FT
FT	VARIANT 604 604
FT	FT
SO	SEQUENCE 651 AA; 72503 MW; 950BABFCBB6A1536 CRC64;
Query Match	100.0%; Score 3326; DB 1; Length 651;
Best Local Similarity	100.0%; Pred. No. 1,3e-224;
Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	

ID	ABG5_MOUSE	STANDARD	PRT	652 AA
Db	1	MDLSLSLTFGGSGMGLVNNRSGSSLSGAPATAPRPHSLGIIMASYSVSHNRVRPMMDITSC	60	
Qy	61	ROOMTQQLIKDVSLLYVESGQIMCITLSSSGSGKTTLLDMSGRIGRACTFLGEYVNGRAL	120	
Db	61	ROOMTQQLIKDVSLLYVESGQIMCITLSSSGSGKTTLLDMSGRIGRACTFLGEYVNGRAL	120	
Qy	121	RREOFQDFCSYVLOSPTLLSLTVRETLNHTALLAIRGPGSGFQKVEVMVMEISLSHV	180	
Db	121	RREOFQDFCSYVLOSPTLLSLTVRETLNHTALLAIRGPGSGFQKVEVMVMEISLSHV	180	
Qy	181	ADRLIGNYSLGISTGERRRVSIAAOLLDPKVMLEFDEPTTGDCWTANOIVLVLELAR	240	
Db	181	ADRLIGNYSLGISTGERRRVSIAAOLLDPKVMLEFDEPTTGDCWTANOIVLVLELAR	240	
Qy	241	RNRIVYLTTHOPRSEFLFQLEDRITALLISFGLITICGTPAEMLDFENCCGYPCEPHSNPDF	300	
Db	241	RNRIVYLTTHOPRSEFLFQLEDRITALLISFGLITICGTPAEMLDFENCCGYPCEPHSNPDF	300	
Qy	301	YMDLTSVDPSQSKRELETSKRQVMIESAAKSAICKTKLTNIEMRKHKTLPVVPKTKD	360	
Db	301	YMDLTSVDPSQSKRELETSKRQVMIESAAKSAICKTKLTNIEMRKHKTLPVVPKTKD	360	
Qy	361	SGGVSKLGLVLLRRTYRNLVYRNKLAVYTRLLQMLGFLLEFVLVYRNSVNLKGAIDRV	420	
Db	361	SGGVSKLGLVLLRRTYRNLVYRNKLAVYTRLLQMLGFLLEFVLVYRNSVNLKGAIDRV	420	
Qy	421	GLLYQVGVATPTTGMNAVNLFPVLAASVDSQSDGLYOKWQMLLYALHVPFSVATM	480	
Db	421	GLLYQVGVATPTTGMNAVNLFPVLAASVDSQSDGLYOKWQMLLYALHVPFSVATM	480	
Qy	481	IFSSVCYMTLGLHPEVYARFGYSALLAARHLIGEFLTVLLGLVQNPNTVNSVVALLSIA	540	
Db	481	IFSSVCYMTLGLHPEVYARFGYSALLAARHLIGEFLTVLLGLVQNPNTVNSVVALLSIA	540	
Qy	541	GLVYSGEFLRNQEMDPIPKIISYTFQKCYCEILVYVNFYGLNPFQSSNNVSYTTPNMC	600	
Db	541	GLVYSGEFLRNQEMDPIPKIISYTFQKCYCEILVYVNFYGLNPFQSSNNVSYTTPNMC	600	
Qy	601	AFTQGIQFTEKTCPGATSRFTNMFLLYSFIPALVILGIVEPKIRDHLSR	651	
Db	601	AFTQGIQFTEKTCPGATSRFTNMFLLYSFIPALVILGIVEPKIRDHLSR	651	

RT mutations in adjacent ABC transporters.":
 RL Science 290:1771-1775(2000).
 CC -1- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.
 CC -1- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
 CC ABCG8 along a pathway regulating dietary-sterol absorption and
 CC excretion (By similarity).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- TISSUE SPECIFICITY: Expressed in the intestine and, at lower
 CC level, in the liver.
 CC -1- INDUCTION: Upregulated by cholesterol feeding. Possibly mediated
 CC by the liver x receptor/retinoid x receptor (LXR/RXR) pathway.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AF312713; AAG53097.1; -
 CC DR MGD: MGI:1351659; Abcg5.
 CC DR InterPro: IPR003593; AAA_ATPase.
 CC DR InterPro: IPR003439; ABC_transporter.
 CC DR Pfam: PF00005; ABC_tran; 1.
 CC DR ProDom: PD000006; ABC_transportr; 1.
 CC DR SMART: SM00382; AAA; 1.
 CC DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 CC DR ATP-binding; Glycoprotein; Transmembrane; Transport.
 CC KM DOMAIN 1 385
 CC FT TRANSMEM 386 406
 CC FT DOMAIN 407 422
 CC FT TRANSMEM 423 443
 CC FT DOMAIN 444 463
 CC FT TRANSMEM 464 484
 CC FT DOMAIN 485 504
 CC FT TRANSMEM 505 525
 CC FT TRANSMEM 526 529
 CC FT DOMAIN 530 550
 CC FT TRANSMEM 551 622
 CC FT DOMAIN 623 643
 CC FT TRANSMEM 644 652
 CC FT NP_BIND 67 94
 CC FT CARBOHYD 410 410
 CC FT CARBOHYD 585 585
 CC FT CARBOHYD 592 592
 CC SEQUENCE 652 AA: 73244 MW: 80CE37ADCC19771E CRC64:
 Query Match 82.3%; Score 2738.5; DB 1; Length 652;
 Best Local Similarity 80.1%; Pred. No. 1.3e-183;
 Matches 522; Conservative 64; Mismatches 65; Indels 1; Gaps 1;

QY 240 RRRRIYVLIHOPRSELFDLPKIALISLNGELIFCGTAPAMDFNDGCPPEHNSPND 299
 DB 241 RRRRIYVLIHOPRSELFDLPKIALISLNGELIFCGTAPAMDFNDGCPPEHNSPND 300
 QY 300 FYWDLTVSTOSKERIEFTSKRWOMESAKYKSAJCHKTLNIEKRLKTLTPMPPEKTK 359
 DB 301 FYWDLTVSTOSKERIEFTSKRWOMESAKYKSAJCHKTLNIEKRLKTLTPMPPEKTK 360
 QY 360 DSEPGSKGLVLRVATRNVLNKLAVITRLLQNLIMGLFLLFVLRVNSNLKGAIDOR 419
 DB 361 DSEPGSKGLVLRVATRNVLNKLAVITRLLQNLIMGLFLLFVLRVNSNLKGAIDOR 420
 QY 420 VGLTLVQVATPYTGMLNVLNPLVLRVNSNLKGAIDOR 479
 DB 421 VGLTLVQVATPYTGMLNVLNPLVLRVNSNLKGAIDOR 480
 QY 480 MIFSSVCYWTGLLHPEVARFGYSALALPHLIGEFLLVLYGIYONPYNVSVALLSI 539
 DB 481 VIFSSVCYWTGLLHPEVARFGYSALALPHLIGEFLLVLYGIYONPYNVSVALLSI 540
 QY 540 AGVLVSGFLRNIOEMPIFKIISYFFQKYSGLVNEFYGLNFTCGSSNVSYTNDP 599
 DB 541 SGILLIGSGFIRNIOEMPIFKIISYFFQKYSGLVNEFYGLNFTCGSSNVSYTNDP 600
 QY 600 CAPTOGIOPTEKPCGATSRFTMNLILXSFIPALVYIGIVFKRDLISR 651
 DB 601 CAITGVOPTEKPCGATSRFTMNLILXSFIPALVYIGIVFKRDLISR 652
 RESULT 3
 ABG5_RAT STANDARD: PRT; 652 AA.
 AC 099PE7;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 GN ABCG5.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxID=10116;
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Small Intestine;
 RX MEDLINE=20578753; PubMed=11138003;
 RA Lee M.-H., Lu K., Hazard S., Yu H., Shulentin S., Hidaka H., Kojima H.,
 RA Allkmetz R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,
 RA Dean M., Patel S.B.;
 RT Identification of a gene, ABCG5, important in the regulation of
 RT dietary cholesterol absorption.";
 RT Nat. Genet. 27:79-83(2001).
 CC -1- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.
 CC -1- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
 CC ABCG8 along a pathway regulating dietary-sterol absorption and
 CC excretion (By similarity).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AF312714; AAG53098.2; -
 CC DR InterPro: IPR003593; AAA_ATPase.

DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran.1
 DR ProDom: PD000006; ABC_transportr.1.
 DR SMART: SM00382; AAA.1.
 DR PROSITE: PS00211; ABC_TRANSPORTER.1.
 DR KMW ATP-binding; Glycoprotein; Transmembrane; Transport.
 FT DOMAIN 1 385 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 386 406 1 (POTENTIAL).
 FT DOMAIN 407 422 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 423 443 2 (POTENTIAL).
 FT DOMAIN 444 463 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 464 484 3 (POTENTIAL).
 FT DOMAIN 485 504 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 505 525 4 (POTENTIAL).
 FT DOMAIN 526 529 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 530 550 5 (POTENTIAL).
 FT DOMAIN 551 624 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 625 645 6 (POTENTIAL).
 FT DOMAIN 646 652 CYTOPLASMIC (POTENTIAL).
 FT NR BIND 87 94 ATP (POTENTIAL).
 FT CAROHD 585 585 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CAROHD 592 592 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 652 AA; 73342 MM; 4042FE2BAB0AD59 CRC64;

Query Match 82.0%; Score 2727.5; DB 1; Length 652;
 Best Local Similarity 79.4%; Pred. No. 7.7e-183;
 Matches 518; Conservative 68; Mismatches 65; Indels 1; Gaps 1;

1 MGDLSTLPGGSMKLVQNRGSSQSSLEGAPATAPAP-NSLGLIHASYSVSHRYPMWDTTS 59
 1 MGELPFLSPGARPHNNRSGSSISESVTSGEHRHSLGVLNVSFVSNNVGPWMTKS 60
 60 CROOITROILKDVSLYESGQIMCIGSSGSGKTTLDAMSGRGRLGACTFLGEYVNGRA 119
 61 COOMDKRILKDVSLYESGQIMCIGSSGSGKTTLDAMSGRGRLGACTFLGEYVNGRA 120
 120 LRROFODCCFYVOSDTLSSLVREHRTALLAHRGPGSFQKVEAVMAELSLH 179
 121 LRROFODCCFYVOSDTLSSLVREHRTALLAHRGPGSFQKVEAVMAELSLH 180
 121 LRROFODCCFYVOSDTLSSLVREHRTALLAHRGPGSFQKVEAVMAELSLH 180
 180 VADRILGNSLIGISGERRRVSIAOQLDOPKVMLEDEPTGLDCAANOIVLVELA 239
 181 VADOMIGNYNGISGERRRVSIAOQLDOPKVMLEDEPTGLDCAANOIVLVELA 240
 240 RRRNIVVITIQPSESEFQFLDKALISFGLIFCGIPAEMLDPPNDGCGYCPESHNFED 299
 241 RRRNIVVITIQPSESEFQFLDKALISFGLIFCGIPAEMLDPPNDGCGYCPESHNFED 300
 300 FYMDLTVYDQSKREIEFTSKRVOMISAYKKSIAICHTLKNIRMHKLKTLPMVPEKTK 359
 301 FYMDLTVYDQSKREIEFTSKRVOMISAYKKSIAICHTLKNIRMHKLKTLPMVPEKTK 360
 360 DSPGVFSLKLVLLRRVTRNLVRLAVITRLLQMLINGLFLFVLVRVSNVLAGAIDR 419
 361 NPPCMFKGLVLLRRVTRNLVRLAVITRLLQMLINGLFLFVLVRVSNVLAGAIDR 420
 420 VGLLYQFVGATPYTGMLNANVLPVLAVSQESQDGLYQKQWMLALALVLPFSVAT 479
 421 VGLLYQFVGATPYTGMLNANVLPVLAVSQESQDGLYQKQWMLALALVLPFSVAT 480
 480 MIFSSVCYMTLGLPEVAFRGYFSAALLAPHLIGEFITLVGLVONPNINVSVALIST 539
 481 VIFSSVCYMTLGLPEVAFRGYFSAALLAPHLIGEFITLVGLVONPNINVSVALIST 540
 540 AGVLVSGFELNIQEMPIPKRIISYTFQKYSCEILVNVNEFYGLNFGSSNVSVTPNM 599
 541 SGLIGSGFINIRIEMPIPKRIISYTFQKYSCEILVNVNEFYGLNFGSSNVSVTPNM 600
 600 CAPFQGIQFTEKTPGATSRFTNMFLLISTPLALVILGIVKTRDHLISR 651
 601 CSMTQGIQFTEKTPGATSRFTNMFLLISTPLALVILGIVKTRDHLISR 652

RESULT 4
 ID ABG8_MOUSE STANDARD; PRT; 673 AA.
 AC Q9DBM0;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 8 (SterolIn-2).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC STRAIN=C57BL/6J; TISSUE=Liver;
 RX MEDLINE=21344600; PubMed=11452359;
 RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
 RA Ose L., Stalenhoef A.F.H., Miettinen T., Bjorkhem I., Bruckert E.,
 RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Silvastava A.K.,
 RA Patel S.B.;
 RA "Two genes that map to the STSL locus cause sitosterolemia: genomic
 RT structure and spectrum of mutations involving sterolIn-1 and
 RL sterolIn-2, encoded by ABG5 and ABG8, respectively.";
 RM Am. J. Hum. Genet. 69:278-290(2001).
 [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC STRAIN=C57BL/6J; TISSUE=Liver;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai I., Shimagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Mikado I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombert P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 [3]
 RP TISSUE SPECIFICITY, AND INDUCTION.
 RX MEDLINE=20553648; PubMed=11099417;
 RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
 RA Kvitlerovich P., Shan B., Barnes R., Hobbs H.H.;
 RT "Accumulation of dietary cholesterol in sitosterolemia caused by
 mutations in adjacent ABC transporters.";
 RL Science 290:1771-1775(2000).
 CC -!- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.
 CC -!- SUBUNIT: May form heterodimers with ABG5 or be tightly coupled to
 CC ABG5 along a pathway regulating dietary-sterol absorption and
 CC excretion (by similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
 CC produced by alternative splicing.
 CC -!- TISSUE SPECIFICITY: Expressed in the intestine and, at lower
 CC level, in the liver.
 CC -!- INDUCTION: Upregulated by cholesterol feeding. Possibly mediated
 CC by the liver X receptor/retinoid X receptor (LXR/RXR) pathway.
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABG8 (WHITE)
 CC SUBFAMILY.

CC -1- CAUTION: Seems to have a defective ATP-binding region.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF324495; AAK84079.1; -
 CC EMBL: AK004871; BAB23630.1; -
 CC MGI: 1914720; Abcg8.
 CC InterPro: IPR003439; ABC_transporter.
 CC Pfam: PF00005; ABC_tran.1.
 CC ProDom: PD000006; ABC_transporter.1.
 CC ProSITE: PS00211; ABC_TRANSPORTER.1.
 CC Glycoprotein; Transmembrane; Transport; Alternative splicing.
 CC DOMAIN 1 413
 CC TRANSMEM 414 434
 CC DOMAIN 435 447
 CC TRANSMEM 448 468
 CC DOMAIN 469 496
 CC TRANSMEM 497 517
 CC DOMAIN 518 526
 CC TRANSMEM 527 547
 CC DOMAIN 548 569
 CC TRANSMEM 570 590
 CC DOMAIN 591 639
 CC TRANSMEM 640 660
 CC DOMAIN 661 673
 CC CAROHD 619 619
 CC VARSPLIC 377 377
 CC SEQUENCE 673 AA: 75995 MW: 7801261.5 Df: 589 CRC64;
 CC
 CC Query Match 21.0%; Score 698; DB 1; Length 673;
 CC Best Local Similarity 28.7%; Pred. No. 3.2e-41;
 CC Matches 194; Conservative 133; Mismatches 264; Indels 84; Gaps 18;
 CC
 CC 11 GSMGVNRKSSQSSLEGAPAPAPPHSGILHASTSVSHVR-----PMD-I 57
 CC 14 GTVLQDASQGLQDSL-----FSESDNSLYFTYSGOSNLEVRDLTYQVDIASQYVFPQL 69
 CC 58 TSCROOQROI-----LKDVSLYVESGOIMCITLSSGSGKTTLLDAMSGRLGRTAF 109
 CC 70 AOFKTPMHSOSSODCEIGINRLSKFVNSGOMLITSSGGRASLDLVITGR-GHGKM 128
 CC 110 -LGEVTVNGALRRQPODFESYVLSQPTLLSLTVRETLHYTALLAI-RGNGPQSFQK 167
 CC 129 KSGQWINGQPTPOLVAKVAVHROHQLPNTLYRETLAFIQMRPRFSAQNRKR 188
 CC 168 VEAVALSLSHVADRLIGNSLGISTGERRRYSIAQLQDPKVMLEDPPTGLDCMT 227
 CC 189 VEDVIAELRLQACNTRVNTYRGVSGGERRRYSIGQILMNGIILLDEPTGLDSEFT 248
 CC 228 ANQIVLVLLVLELARRRIYVLIHQPRSELPLQLEKIALILSGELIFCOTPRMELDFDC 287
 CC 249 AHNVTTLRLAKGNRLVLSLHOPRSDFPLFLVLLMNTSGTPIYLSAQAOMQYFSLI 308
 CC 288 GYCPPEHSNPEDFYMDLTSVDQSKEREIETSKRVOMIESAYKRS-----ICHKTLKN 341
 CC 309 GHCPKRYSPADFYVDLTSIDRSKEREVAVEKAQSLALFLKQYGFDFLKAKEKE 368
 CC 342 IERKHKLTLPVVFETKDS-----PGVSKLGVLLRRVTRNLVKNKLAIVITRLQN 393
 CC 369 LMTSTHTVSLTL-----TQDTCGTAVELPGMIEQFSLIRQISNDRDPLTLIHSEA 424
 CC 394 LIMGLFLFLFYLRVSNVTKGALQDRVGLLYQFQCAPPYGMNAVULFPLRAVSDQES 453
 CC 425 CIASLIIGFLYTGGAQOL--SPMDTALFLMIGALIPFNVTIDVVSCKHSESRMLYEL 482
 CC 454 QDGLYQWOMMLAVLHVLFPFSVATMIFSSVCWTGLHPEVARFQFYSAALLAPHLIG 513

DB 483 EQLTAGYFFFAKILGELPEHCAYIITAMPIYLTNRPPVEL-----LL--HELL 534
 OY 514 EFL-----TVLLGIVONPNI-VNSVALLSTAGLVGSGFLRNIOEMPIKIISYFT 566
 DB 535 VMLVVECCRTMALAASAMLPFTFMSFCNALYNSFYLAGFMINDMLIYPAWTSKLS 594
 OY 567 FQKYSCEIIVNVEFGLNFT--CGSSNVSTTPKCAFTQGIQFIKTCRGATSRFTNMF 624
 DB 595 FLKWCSSGLMOIQFONCHLVTYTOIGNFTFSLIGDTM-----ISANDLMS 637
 OY 625 LILYSFIPALVILGI 639
 DB 638 HPLXALY--LIVIGI 650
 RESULT 5
 ABG8_HUMAN STANDARD; PRT; 673 AA.
 ID ABG8_HUMAN
 AC 09H21:
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 8 (Sterolin-2).
 GN ABG8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 OX [1]
 RP SEQUENCE FROM N.A., VARIANTS STIOSTEROLEMIA T-231; Q-263; R-574 AND
 RP R-596, AND VARIANT C-54.
 RX MEDLINE=2053648; PubMed=11099417;
 RA Berge R.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
 RA Kwaterovich P., Shan B., Barnes R., Hobbs H.H.;
 RT "Accumulation of dietary cholesterol in stiosterolemia caused by
 RT mutations in adjacent ABC transporters.";
 RL Science 290:1771-1775(2000).
 RN [2]
 RN SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), VARIANTS STIOSTEROLEMIA H-184;
 RN T-231; Q-263; H-405; P-501; S-543; P-572; E-574; R-574; R-596 AND
 RN F-570 DEL, AND VARIANTS H-19; C-54; K-238; V-259; K-400; R-575 AND
 RN A-632.
 RC TISSUE=Liver;
 RX MEDLINE=21344600; PubMed=11452359;
 RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hida H., Kojima H.,
 RA Ose L., Stalenhoef A.F.H., Miettinen T., Bjorkhem I., Bruckert E.,
 RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,
 RA Patel S.B.;
 RT "Two genes that map to the STSL locus cause stiosterolemia: genomic
 RT structure and spectrum of mutations involving sterolin-1 and
 RL Am. J. Hum. Genet. 69:278-290(2001).
 RN [3]
 RN REVIEW.
 RX MEDLINE=21474438; PubMed=11590207;
 RA Schmitz G., Langmann T., Helmerl S.;
 RT "Role of ABG1 and other ABCG family members in lipid metabolism.";
 RL J. Lipid Res. 42:1513-1520(2001).
 CC -1- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.
 CC -1- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to
 CC ABCG5 along a pathway regulating dietary-sterol absorption and
 CC excretion.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: 1 (shown here) and 2; are
 CC produced by alternative splicing. Isoform 2 is a minor form
 CC detected in approximately 10% of the cDNA clones.
 CC -1- TISSUE SPECIFICITY: Strongly expressed in the liver, lower levels
 CC in the small intestine and colon. Detectable in a wide variety of
 CC human tissues.
 CC -1- DISEASE: Defects in ABG8 are a cause of stiosterolemia, also

RESULT 6
ABCG8_RAT
ID ABCG8_RAT STANDARD: PRT: 672 AA.
AC p58428;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 8 (Sterolin-2).
GN ABCG8.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN=Sprague-Dawley;
RX MEDLINE=21344600; PubMed=11452359;
RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H., Ose U., Stalenhoef A.F.H., Metlinen T., Bjorkhem I., Bruckert E., Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K., Patel S.B.;
RA "Two genes that map to the SRS1 locus cause sitosterolemia: genomic structure and spectrum of mutations involving sterolin-1 and sterolin-2, encoded by ABCG5 and ABCG8, respectively.*";
RT Am. J. Hum. Genet. 69:278-290(2001).
RT -1- FUNCTION: Transporter that appears to play an indispensable role in the selective transport of the dietary cholesterol in and out of the enterocytes and in the selective sterol excretion by the liver into bile.
CC -1- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to ABCG5 along a pathway regulating dietary-sterol absorption and excretion (by similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be produced by alternative splicing.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE) SUBFAMILY.
CC -1- CAUTION: Seems to have a defective ATP-binding region.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF351785; AAK84831.1; -
DR InterPro: IPR003593; AAA_Arase.
DR pfam: PF00005; ABC_transpor.
DR SMART: SM00383; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW Glycoprotein; Transmembrane; Transport; Alternative splicing.
FT DOMAIN 1 412
FT TRANSMEM 413 433
FT DOMAIN 434 446
FT TRANSMEM 447 467
FT DOMAIN 468 495
FT TRANSMEM 496 516
FT DOMAIN 517 525
FT TRANSMEM 526 546
FT DOMAIN 547 568
FT TRANSMEM 569 589
FT DOMAIN 590 628
FT TRANSMEM 629 649
FT DOMAIN 650 672
FT CARBOHYD 618 618
FT VARSPIC 376 376
SQ SEQUENCE 672 AA: 75878 MW: 6088973F3FA4C36F CRC64;
Query Match 20.8%; Score 691; DB 1; Length 672;
Best Local Similarity 28.3%; Pred. No. 9.8e-41;

Matches 169; Conservative 126; Mismatches 264; Indels 88; Gaps 15;
QY 23 SSLEGAAPANAPEPHSLGILHASYSVSHRYR-----PW-----MDITSC 60
DB 21 SSLQDSVFSESDNSLYFTYSQGSNTLEVRDLYQYDMAQVPMFQALQFKLPMRSRG 80
QY 61 ROOMTOILKDVSLVYESQIMKILSSGSGKTTLLDMSGRGLRGRTGLGEVYVNGRAL 120
DB 81 QDSMDLGI-RNLSFKVRSQMLAIGSAGCGRATLLDVTGTRHGCKMKSGQIWMGOPS 139
QY 121 RREOFQDCFSYVQSDTLLSLVRETLHTALAIRGNPGSF-----QKKEVMAEL 175
DB 140 TQQLDLCYAHVAGQDQLPMLNLTVELTFTIAQRL-----PKTESQAQRDRVEDYIAEL 195
QY 176 SLSHVADRLLIGNSLGISTGERRRYSIAAQLDPRKVMLEPTTGLDQMTANQIVLL 235
DB 196 RLKOCANTRVGNTRYGVSGGERRRYSIGVQLLMPNGIILDEPTSGDLSFTAHNLVRL 255
QY 236 VELARRNRIVLTIRHPRSELFOLEPKIALISGELIFCGTPAEMLDENDCGPPERS 295
DB 256 SRLAKGNRLVLTSLHPRSDIFRLFDVLVLTMTSGTPIYLGVAOHMVOYFTSIGYPCPRS 315
QY 296 NPFDFYMDLTSVDTSQEREIEFSKRVOMIESAYKKS-----ICHKTLKNIEMKHLK 349
DB 316 NPADFVDTSLDRSKEDQVATMERARLLALFLERKVGDFDEFLKMAEKSID-----TG 371
QY 350 TLPMPPEKTKDS-----PGVFSKIGVLRVTRNLYRANKLAVITRLQLNLMGLFTL 401
DB 372 TYAVSQTLNQDNFCGAELPGLIQFTLLIRNOISNDRDLPTLTHGAENCLMLIIG 431
QY 402 FVYLKRSVNLKGAIDRGLYQFYGAFTYGMALNVLFPYLRAVSQDSQGLYQYM 461
DB 432 FLYYGHADRL--SFMDMAALLFMIGALIPFNWILLDVVSKCHERSILYLEDGLYTG 489
QY 462 QMWLYALHLPESVYATMIFSSCYWTGLHPEVARPFYSALLAPLIGEL-----516
DB 490 PYEFKAVIGELDEHCAVYITIGMPYITWLNLR-----GPEFLHFMILMLVFPCC 541
QY 517 -TLVLGLIVQNPNV-VNSVALLSLAGVYVSGFLRNIOEMPIFEKIIISYTFQKCSSEI 574
DB 542 RTMALAASAMLPFTFHMSSCCNALNVSFTLTFGFMNLNMLIVPMWIKMFLKRCFSG 601
QY 575 LVVNEFYGLNFCGSSNVSYTNPMCAFTQGIQIEKTCPP--ASRFTANFLIYSFIP 632
DB 602 IMQIOFNHITYTTOIGNLTFV-----PGDAMVTAMDLSNPILYAIY- 643
QY 633 ALVIIGI 639
DB 644 -LIVIGI 649
RESULT 7
ABG2_HUMAN
ID ABG2_HUMAN STANDARD: PRT: 655 AA.
AC Q9DNO0; Q95374; Q9NDS0; Q9BY73;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 2 (Placenta-specific ATP-binding cassette transporter) (Breast cancer resistance protein).
GN ABG2 OR ABCP OR BCRP OR BCRP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=9906513; PubMed=9850061;
RA Allikmets R., Schmitt L.M., Hutchinson A., Romano-Spica V., Dean M.;
RT "A human placenta-specific ATP-binding cassette gene (ABCG) on chromosome 4q22 that is involved in multidrug resistance.";
RL Cancer Res. 58:5337-5339(1998).

RN [12] SEQUENCE FROM N.A.
 RP TISSUE-Breast cancer;
 RC MEDLINE-99080071; PubMed-9861027;
 RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,
 RA Ross D.D.;
 RT "A multidrug resistance transporter from human MCF-7 breast cancer
 cells";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:15665-15670(1998).
 RN [13]
 RP ERRATUM.
 RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,
 RA Ross D.D.;
 RL Proc. Natl. Acad. Sci. U.S.A. 96:2569-2569(1999).
 RN [14]
 RP SEQUENCE FROM N.A.
 RA Kage K., Tsukahara S., Sugiyama T., Asada S., Ishikawa E., Tsuruo T.,
 RA Sugimoto Y.;
 RT "Breast cancer resistance protein constitutes a 140-kDa complex as a
 homodimer";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RN [15]
 RP SEQUENCE OF 198-655 FROM N.A.
 RC TISSUE-Placenta;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Matanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Makamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
 RA Niinomiya K., Iwayanagi T.;
 RT "NDO human cDNA sequencing project";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [16]
 RP REVIEW.
 RA MEDLINE-21474438; PubMed-11590207;
 RA Schmitz G., Langmann T., Heimerl S.;
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism";
 RL J. Lipid Res. 42:1513-1520(2001).
 CC -1- FUNCTION: XENOBIOTIC TRANSPORTER THAT APPEARS TO PLAY A MAJOR ROLE
 IN THE MULTIDRUG RESISTANCE PHENOTYPE OF A SPECIFIC MCF-7 BREAST
 CANCER CELL LINE. WHEN OVEREXPRESSED, THE TRANSFECTED CELLS BECOME
 RESISTANT TO MITOXANTHONE, DAUNORUBICIN AND DOXORUBICIN. DISPLAY
 DIMINISHED INTRACELLULAR ACCUMULATION OF DAUNORUBICIN, AND
 MANIFEST AN ATP-DEPENDENT INCREASE IN THE EFFLUX OF RHODAMINE 123.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF103796; AAD09188.1;
 DR EMBL: AF098951; AAC97367.1;
 DR EMBL: AB056867; BAB39212.1;
 DR EMBL: AK002040; BAA92050.1;
 DR Genev: HGNC:74; ABCG2.
 DR MIM: 603756;
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transpfr.
 DR Pfam: PF00005; ABC_tran.1.
 DR Prodom: PD000006; ABC_transpfr.1.
 DR SMART: SM00382; AAA.1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; FALSE NEG.
 KW ATP-binding; transmembrane; transport.
 FT DOMAIN 1 395 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 396 416 POTENTIAL.
 FT DOMAIN 417 428 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 429 449 POTENTIAL.
 FT DOMAIN 450 477 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 478 498 POTENTIAL.
 FT DOMAIN 499 506 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 507 527 POTENTIAL.
 FT DOMAIN 528 535 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 536 556 POTENTIAL.
 FT DOMAIN 557 630 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 631 651 POTENTIAL.
 FT DOMAIN 652 655 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 80 87 ATP (POTENTIAL).
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 596 596 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 24 24 V -> A (IN REF. 2 AND 4).
 FT CONFLICT 166 166 E -> Q (IN REF. 2 AND 4).
 FT CONFLICT 208 208 F -> S (IN REF. 1).
 FT CONFLICT 315 316 MISSING (IN REF. 5).
 FT CONFLICT 482 482 R -> T (IN REF. 2).
 SO SEQUENCE 655 AA; 72343 MM; 89A6D351IDSCCE0 CRC64;
 Query Match 20.3%; Score 676.5; DB 1; Length 655;
 Best Local Similarity 29.0%; Pred. No. 9.8e-40;
 Matches 181; Conservative 137; Mismatches 251; Indels 55; Gaps 18;
 QY 21 SSSLEGAPATAP---EPHSIGILHASYSVSHRPMPMDITSCROQWTRQIKQSYLVE 77
 DB 13 SGGNTNGEPRAVSDNKAFTGCAVLSFNITCYRKAKSGFLPCKPVEKELTSLNGIMK 72
 QY 78 SGQIMCLIGSSGSKRTLLDMSGRGLAGAGTFLGEVYVNGRALRREDPODFSVYASDT 137
 DB 73 PG-LNAILGPTGGCKSSLDVLAARKDPSG-LSGDIVLNG-APPANPKNSGYVODV 129
 QY 138 LLSLTVRETHYRALLAIRGNGC-SQKQVEAVMAELSHVADRLIGYSLAGISTG 196
 DB 130 VGLTVRENLQFSAAKLRLATMTNHEKNERINIVIELGDKVADSVMGTQFTRGVSGG 189
 QY 197 ERRRVSTAAQLLDOPKYALFDEPTTGLDCAANOIVYLVELARRNRIVYLTHIOPSEL 256
 DB 190 ERKRTISGMELTTPSLIFLDEPTTGLDSTANNAVLLILKMSKQGRITLFSHQPRYSI 249
 QY 257 FQLPDKTALLSFGELIFCGPAEMLDFFNDGYPCEPHSNPFYMDLTSDVDTQ-----SK 312
 DB 250 FKLPDSLTLLASGLMFGHPAQELAGFEASGYCEAVNNPADFLDINDSDVAVALNR 309
 QY 313 ERE-----LETSR-----VOMISAKKSKALCKT-----LNIERMKHLKTLPMVPF 356
 DB 310 EEDFKATELLEPSKQDKPLIEKLAETIYVNSSEFYETRAEDHQLSGEKKKITYFKEISY 369
 QY 357 KTKDSPGVFSKGLVLRVTRNLRNKLAVITRLLQMLINGFLF-LFVLRVRSNYLKG 414
 DB 370 TT-----SFHQQLRNVRSKRSKNLGNQASIAQIIVYVGLVIGALIFGLKNDST----- 421
 QY 415 AIQDRVGLLYQFGATPYTGLNLANVNLFPYLRAVSDQSDGLYQKQMMALVAL-HYLP 473
 DB 422 GIONRAVLFR-FLTTNOCFSSVSVAVELFVVEKKLFIEHYISGYRVSSYFLGRLSLDLP 480
 QY 474 FSVYATMIFSSVCWTGLGHPVARFCYFSAALLAPHLIEFLTVLLGIYQNNIYNSV 533
 DB 481 MRMLPSIIFTCIYIFMLGLPKADAFVMMFTLL--MVAYSASMAALAAAGOSVYSA 537
 QY 534 VALSLAGV--LVSGFLRNIOEMPIFKIISYFTFOKYCEILIVVEEFGNLGTCSSN 591
 DB 538 TLMTICFVPMIIFSGILVALLTITIASLSLQIFSIPIRYGFTALQNHNEFGQNFPCG--- 594
 QY 592 VSVYTPNMCAPFTGQIOFIKTCG 615
 DB 595 LNAATGNPCMYA-----TCYG 610
 RESULT 8
 YOH5_YEAST
 ID YOH5_YEAST STANDARD; PRT; 1294 AA.

AC 008234; 008233;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable ATP-dependent transporter YOL074C/YOL075C.
 GN YOL074C/YOL075C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxId=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-97321807; PubMed-9178509;
 RA Tzeremia M., Katsoulou C., Alexandraki D.;
 RT Sequence analysis of a 33.2 kb segment from the left arm of yeast
 RT chromosome XV reveals eight known genes and ten new open reading
 RT frames including homologues of ABC transporters, histidinol
 RT phosphatases and human expressed sequence tags.*;
 RL Yeast 13:383-389(1997).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY MDR SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

 DR EMBL: 274817; CAA99085.1;
 DR EMBL: 274816; CAA99084.1;
 DR SGD: S0005435; YOL075C.
 DR InterPro: IPR003593; AAA_Arpase.
 DR InterPro: IPR003439; ABC_Transport.
 DR Pfam: PF00005; ABC_tran; 2.
 DR ProDom: PD000006; ABC_transport; 1.
 DR SMART: SM00382; AAA; 2.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
 DR Hypothetical protein; ATP-binding; Transmembrane; Glycoprotein;
 KW Transport; Repeat.
 FT TRANSMEM 376 396 POTENTIAL.
 FT TRANSMEM 496 516 POTENTIAL.
 FT TRANSMEM 531 551 POTENTIAL.
 FT TRANSMEM 605 625 POTENTIAL.
 FT TRANSMEM 1039 1059 POTENTIAL.
 FT TRANSMEM 1121 1141 POTENTIAL.
 FT TRANSMEM 1267 1287 POTENTIAL.
 FT NP_BIND 62 69 ATP (POTENTIAL).
 FT NP_BIND 727 734 ATP (POTENTIAL).
 FT CARBOHYD 41 41 N-LINKED (GLCNAc...) (POTENTIAL).
 FT CARBOHYD 86 86 N-LINKED (GLCNAc...) (POTENTIAL).
 FT CARBOHYD 101 101 N-LINKED (GLCNAc...) (POTENTIAL).
 FT CARBOHYD 151 151 N-LINKED (GLCNAc...) (POTENTIAL).
 FT CARBOHYD 341 341 N-LINKED (GLCNAc...) (POTENTIAL).
 FT CARBOHYD 349 349 N-LINKED (GLCNAc...) (POTENTIAL).
 FT CARBOHYD 371 371 N-LINKED (GLCNAc...) (POTENTIAL).
 FT CARBOHYD 528 528 N-LINKED (GLCNAc...) (POTENTIAL).
 FT CARBOHYD 983 983 N-LINKED (GLCNAc...) (POTENTIAL).
 FT CARBOHYD 1062 1062 N-LINKED (GLCNAc...) (POTENTIAL).
 SO SEQUENCE 1294 AA; 145157 MW; C555500A45E9284E CRC64;

Query Match 18.9%; Score 627; DB 1; Length 1294;
 Best Local Similarity 31.7%; Pred. No. 6,7e-36;
 Matches 181; Conservative 106; Mismatches 228; Indels 56; Gaps 19;

OY 65 TROILKDVSLVSESGQICLSSGSGKTTLLDANGSLRAGTFL-----GEVYVNRRA 119
 DB 706 TREILQSVNAIFKRGIMVIMPGSSGSSLLNLISGL KSSFAKEDTSGSIFNDIQ 764
 OY 120 LRROPODCSYVLSQST-LISLTVRETLHYTALLAIRGNPGSFORKVAVNAELSL 178
 DB 765 VSELIFKVCYSVSDDDHLALATLVETLKYAALRLAHULTEARBERTNLIRSLG 824

OY 179 HVADRILGNVSLGISTGERRRYSIAQALODPKWMLFDEPTGLDCMTANQVLLVEL 238
 DB 825 HCENNITIGNEFVNGISGGERRRYTMGOVLLNDPILLDEPSSGLSFTSATILELEKL 884
 OY 239 AR-RNRIVULTIHOPESELFOLFDKTAIIS-FCGLJFCGPPAMLPFNDCCGRCDEHN 296
 DB 885 CROGQTIITTHQPSSELFKRGVNLAKSRTPANSPDEMIATFELGNCPSFTN 944
 OY 297 PEDFYMDLPSVDTSKREIETSKRYQMESAYKSAICHTLKINTERKMLTLPMPVP 356
 DB 945 VADFFDLISVTQNBQNEISSRAVEKILISANK-----ANNDN-ESLSPTISER 994
 OY 357 KTRDSGVSKGLVLLRRTNRY-----RNLAVITRLQNLIMGLFLL 401
 DB 995 OQYOSSEFTEYSEFYRK-PANLVLAIVNVKROPTTTRRSFDSLMARIAQJLGLVIRA 1053
 OY 402 FEVLVRSNVLKGAIDRGVGLYQFVGATPYGMVAVNLFPYLRVAVSQESQDGLYOKM 461
 DB 1054 LEFAPVKHNT--SISNRLGLQEST-ALIEVGMGLNLCYPERDYEEYENDVYGLA 1110
 OY 462 QMMLAYALHLPESVAVATMIFSSVCYTGLLPEVARFEGFSALAPHLI--GEFLTL 518
 DB 1111 PEFLLAVMTLEPLSLASVLAIVFVYLAGL-PRTA--GNFATVYCSPIVCCGRRLG 1167
 OY 519 VLLGIYQNN-IVNSVVALSLAGVYSGFLRNIDEMPIPKIISYFTFOKYSILVY 577
 DB 1168 MNTFEFERGFVANCISIIISIGTQMSGMLST-----GMSRVLKGFVNLVPVGTSMIIN 1223
 OY 578 NEFYG-LNFTC--GSSNVSVTNPMAFTQG 605
 DB 1224 FAPPGMLKLTCEGDKNSDGT-----CEPANG 1250

RESULT 9
 WHIT_LUCCU
 ID WHIT_LUCCU STANDARD; PRT; 677 AA.
 AC 005360;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE White protein.
 GN Lucilia cuprina (Greenbottle fly) (Australian sheep blowfly).
 OS Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;
 OC Muscomorpha; Oestroidea; Calliphoridae; Lucilia.
 OX NCBI_TaxId=7375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-97087158; PubMed-89333176;
 RA Garcia R.L., Perkins H.D., Howells A.J.;
 RT "The structure, sequence and developmental pattern of expression of
 RT the white gene in the blowfly *Lucilia cuprina*.";
 RL Insect Mol. Biol. 5:251-260(1996).
 SO SEQUENCE OF 490-584 FROM N.A.
 MEDLINE-90264941; PubMed-1971656;
 RA Elizur A., Vacek A.T., Howells A.J.;
 RT "Cloning and characterization of the white and topaz eye color genes
 RT from the sheep blowfly *Lucilia cuprina*.";
 J. Mol. Evol. 30:347-358(1990).
 CC -1- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
 CC NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
 CC CELLS RESPONSIBLE FOR EYE COLOR.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY MDR SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

RX MEDLINE-96423158; PubMed-8825759.
RA Beauseny N.J., Bedell J.A., Benedict M.O., Mukabayire O., Hliliker D.,
RT Collins F.H.;
CL "Cloning and characterization of the white gene from Anopheles
gambiae.";
RL Insect Mol. Biol. 4:217-231(1995).
CC -1- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
CELLS RESPONSIBLE FOR EYE COLOR.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.ch/announce/or send an email to license@isb-sib.ch).

CC CC-----
DR EMBL_029486; AAC46995.1; .-
DR EMBL_029485; AAC46994.1; .-
DR EMBL_029484; AAC47423.1; .-
DR InterPro: IPR003593; AAA_Alpase.
DR InterPro: IPR003439; ABC_transpotr.
DR InterPro: IPR005284; Pigment_permease.
DR Pfam: PF00005; Abc_tran; 1.
DR ProDom: PD000006; ABC_transportr; 1.
DR SMART: SM00382; AAA; 1.
DR TIGRFAMS: TIGR00955; 3a01204; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
DR Pigment; ATP-binding; Transmembrane; transport.
KW NP_BIND 133 140
FT FT 288 295 ATP (POTENTIAL).
FT FT TRANSMEM 444 464 ATP (POTENTIAL).
FT FT TRANSMEM 474 494 POTENTIAL.
FT FT TRANSMEM 524 544 POTENTIAL.
FT FT TRANSMEM 552 572 POTENTIAL.
FT FT TRANSMEM 581 601 POTENTIAL.
FT FT CARBOHYD 472 472 POTENTIAL.
FT FT CARBOHYD 645 645 N-LINKED (GLCNAC...) (POTENTIAL).
FT FT CONFLICT 100 100 N-> S (IN REF. 1; AAC47423) .
FT FT CONFLICT 691 693 SRS -> Y (IN REF. 1; AAC47423) .

SQ SEQUENCE 695 AA; 77218 MW; EE8B9517239B2961 CRC64;

Query Match 18.3%; Score 607.5; DB 1; Length 695;
Best Local Similarity 28.4%; Pred. No. 6; ge-35;
Matches 170; Conservative 124; Mismatches 208; Indels 97; Gaps 17;

QY 58 TSCRQQWTRQ-----TLKDVSLVYESGQIMCIIIGSSGGSKTKLLDMSGRLRAGCT 108
Db | ||| :||:|: |:||||:|||||||::|: |
96 TLRLNCRSTRKRDPNRPKHLIKNTGYAKSGELIAYVGSSGAQKTLLNLALRR-SPPGV 154

QY 109 FLGEVVY---NGRALRREDFQDFCSVYLASDTLSLTSLIVARETHLYATALLAIRGNCGSF- 164
Db | ||| :||:|: |:||||:|||||||::|: |
155 KISPNAVRALNGVPVAHQEARCAVCAYVOODDLFPISLTRRHLLFOAMLRMGDVASRYK 214

QY 165 QKVVEAVMAFLSTSHVADRLIGNYS-LGIISTGERRVASTIAOQLLODKPVMLFDPEPTGL 223
Db | :||:|: ||||| :||:|: |:||||:|||||||::|: |
215 QHRVQEVLDELTVMKCADVTILAPGRKIKGSGGERKLAFASFETLDPHILLDCDEPTSL 274

QY 224 DCATANQIVLVLELARBNRIYVLTTHORPSLFOLFEDITALISPGELLFCGPAMIDPF 283
Db | :||:|: |:||||:|||||||::|: |
275 DSMANSAVLAOVLMGMKMGTITLTITHOSSELYLCLEDKILVAEGSRVAFILSPYQSABE 334

QY 284 FNDCGYPCEHNSNPFDYMDLTSVDPOQRERELETJSRKVMIESAYKKKAICKTKLNIE 343
Db | ||| :|||:|: ||||:|: |:||||:|||||||::|: |
335 FSQGLGPCRPNVPAPOFYVOLMI--APAKEAECDMIIKKICDSRAVSPIAREVLETFS 391

QY 344 RKNHKLTLTPVHPFKTDSPGVSKL-GV-----LLRRVRTNLRNKLIAYI 387
Db -----VAGKGDEPYMTAQGVESGTYGSRSMTQFFCIILRSMISLVLKIDPLVK 441

OY 388 TRLLONLIMGLFLEFVLKNSVNLKAIODRVGL-----LYOFVGATPYTGMLNAVNL 441
 DB 442 VALLQTFANVATLI-----GSIYFGVLDQDGMVNINGSLFLFLTNFTQNVAVINV 493
 OY 442 FVVLKAVSDQESQDGLXQKQWQMLVAVLHVPFVSVAWIMFSSVCYTWLGLHPEARVRCY 501
 DB 494 FSAELPVLKREKRSRIYRDYFLKCTIAELPLFAVPFVFSIYPMIGL-----RTG- 547
 OY 502 FSAALLAPHLGEFLTLVLGIVQNP-----IVNSVVALSLIA-----GYLVCSSG 547
 DB 548 -----ATHYL-----TTLFVTLVAVNSFSLGLICASSISMSALSVCPVYIPLIFGG 598
 OY 548 FLRNIOEPIPKIISYTFYFQKCEILVNEFYGL-----NFTCGSSNVSVYT 596
 DB 599 FLNLSAVPAYFKYLSYLSMFYRANEDALLINQMSVYVQELACTRANTVCPREIILET 657

RESULT 12

WHIT_DROME

ID WHIT_DROME STANDARD: PRT: 687 AA.

AC P10090: Q9Y3A2: Q9Y333;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-NOV-1991 (Rel. 20, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE White protein.

GN W OR EG: BACN33B1.1 OR CG2759.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

OC NCBI-TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-Head;

RX MEDLINE=90221897; PubMed=2109311;

RA Repling M., Mount S.M.,
 "Sequence of a cDNA from the Drosophila melanogaster white gene.";

RL Nucleic Acids Res. 18:1633-1633(1990).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=85134865; PubMed=6084717;

RA O'Hare K., Murphy C., Lewis R., Rubin G.M.,
 "DNA sequence of the white locus of Drosophila melanogaster.";

RL J. Mol. Biol. 180:437-455(1984).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=21100348; PubMed=11156992;

RA Lukatsovich T., Asztalos Z., Awano W., Baba K., Kondo S., Miwa S.,
 Yamamoto D.,
 "Dual-tagging gene trap of novel genes in Drosophila melanogaster.";

RL Genetics 157:727-742(2001).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN-Berkeley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 Brandon R.C., Rogers Y.H.C., Blaise R.G., Zhang C., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Beeson K.Y., Benos P.V., Bernan B.P., Bhandaril D., Bolshakov S.,
 Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottlter P.,
 Burdits K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
 Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mervulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Mobarry C., Morris J., Moshrefi A.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paoleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of Drosophila melanogaster.";

Science 287:2185-2195(2000).

[5]

RP SEQUENCE FROM N.A.

RC STRAIN-Oregon-R;

RX MEDLINE=20196011; PubMed=10731137;

RA Benos P.V., Gatt M.R., Ashburner M., Murphy L., Harris D.,

RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demallies J., Cadieu E.,

RA Dreano S., Gloux S., Leilaure V., Mottier S., Galibert F., Borkova D.,

RA Minna B., Kalatos F.C., Louis C., Siden-Klamos I., Bolshakov S.,

RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,

RA Modolell J., Peter A., Schoettler P., Werner M., Moulfict F.,

RA Belnert N., Dove G., Schaefer U., Jackle H., Bucheton A.,

RA Callister D.M., Campbell L.A., Parlantous A., Henderson N.S.,

RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,

RA Glover D.M.,

RT "From sequence to chromosome: the tip of the X chromosome of D.

RT melanogaster.";

RL Science 287:2220-2222(2000).

[6]

RP SEQUENCE OF 224-331 FROM N.A.

RX MEDLINE=89339145; PubMed=2503416;

RA Tearle R.G., Belote J.M., McKown M., Baker B.S., Howells A.J.,

RT "Cloning and characterization of the scarlet gene of Drosophila

RT melanogaster.";

RL Genetics 122:595-606(1989).

CC -! FUNCTION: PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM NECESSARY

CC FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT CELLS

CC RESPONSIBLE FOR EYE COLOR. WHITE DIMERIZE WITH BROWN FOR THE

CC TRANSPORT OF GUANINE AND WITH SCARLET FOR THE TRANSPORT OF

CC TRYPTOPHAN.

CC -! SUBUNIT: HETERODIMER OF WHITE WITH EITHER BROWN OR SCARLET.

CC -! SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC

DR EMBL: X51749; CAA36038.1; -

DR EMBL: X02974; CAA26716.1; -

DR EMBL: AB028139; BAA78210.1; -

DR EMBL: AE003425; AAF45826.1; -

DR EMBL: AL133506; CAB65847.1; -

DR PIR: S07263; CAA53795.1; -

DR PIR: S08635; EYFEM.

DR FlyBase: FBgn0003996; w.

DR InterPro: IPR003593; AAA_ATPase.

DR InterPro: IPR003439; ABC_transporter.

Query Match	18.1%	Score 602.5;	DB 1;	Length 687;
Best Local Similarity	28.8%	Pred. No. 1.5e-34;		
Matches 180;	Conservative 11;	Mismatches 220;	Indels 95;	Gaps 19

[illegible]

ID ABG1_HUMAN STANDARD; PRT; 678 AA.
 AC P45844.4, G9BX10, G9BX11, G9BX12, G9BX13, G9BX14, G9BXK6, G9BXK7,
 AC G9BXK8, G9BXK9.
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 1 (White protein homolog)
 DE (ATP-binding cassette transporter 8).
 GN ABCG1 OR ABC8 OR WHT1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE OF 3-678 FROM N.A. (ISOFORMS 1 AND 4).
 RC TISSUE=Retina;
 RX MEDLINE=96256850; PubMed=8659545;
 RA Chen H.M., Rossler C., Ialioi M.D., Lynn A., Chakravarti A.,
 RA Perrin G., Antonarakis S.E.;
 RT "Cloning of the cDNA for a human homologue of the Drosophila white
 RT gene and mapping to chromosome 21q42.3.";
 RL Am. J. Hum. Genet. 59:66-75(1996).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
 RA Ohmi M., Takagi T., Sakaki Y., Taudien S., Bleichschmidt K., Polley A.,
 RA Menzel U., Delabar J., Kumpf K., Lehman R., Patterson D.,
 RA Reichwald K., Rump A., Schillabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Negamine K., Mitsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordstiek G., Hornischer K., Brandt P.,
 RA Scharte M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
 RA Ramser J., Beck A., Klages S., Hennig S., Rlessmann L., Dagand E.,
 RA Weimayer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Lehrach H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=20408883; PubMed=10950923;
 RA Berry A., Scott H.S., Kudoh J., Talior I., Korostilshvsky M.,
 RA Mattenoler M., Guipponi M., Barras C., Rossler C., Shibuya K.,
 RA Wang J., Kawasaki S.E., Asakawa S., Minoshima S., Shimizu N.,
 RA Antonarakis S.E., Bonne-Tamir B.;
 RT "Refined localisation of autosomal recessive nonsyndromic deafness
 RT DNB10 locus using 34 novel microsatellite markers, genomic
 RT structure, and exclusion of six known genes in the region.";
 RL Genomics 68:22-29(2000).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=21192304; PubMed=11127901;
 RA Poroch-Oezueremes M., Langmann T., Helmerl S., Borsukova H.,
 RA Kaminski W.E., Drobnik W., Honer C., Schumacher C., Schmitz G.,
 RA "The zinc finger protein 202 (ZNF202) is a transcriptional repressor
 RT of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene
 RT expression and a modulator of cellular lipid efflux.";
 RL J. Biol. Chem. 276:12427-12433(2001).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 2, 3, 4, 5, 6 AND 7).
 RX MEDLINE=21092576; PubMed=11162488;
 RA Lorkowski S., Rust S., Engel T., Jung E., Tegelkamp K., Galinski E.A.,
 RA Assmann G., Cullen P.;
 RT "Genomic structure and structure of the human ABCG1 (ABC8) gene.";
 RL Biochem. Biophys. Res. Commun. 280:121-131(2001).
 RN [6]
 RP SEQUENCE OF 33-678 FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=97186700; PubMed=9034316;
 RA Croop J.M., Tiller G.E., Fletcher J.A., Lux M.L., Raab E.,
 RA Goldenson D., Archiniegas S., Son D., Wu R.;
 RT "Isolation and characterization of a mammalian homolog of the

Db 248 DSASCFQVYSLMKGLAOGSIICITIHOPSAKLEFEDOLVYLSGOCVYRGKVCNLVY 307
 QY 284 FNDGYPQPEHNSPDEYMDLTSDVTSKEREIETSKRYOMIESAYKSA-----I 334
 Db 308 LRDGLMCPYHNADVMEVASEGYDDONSRLVAVRECKDCSDHRDLGGDAEVNPF 367
 QY 335 CHKTAKNIEMKHLKTLPMVPEKTKDSPGV-----FSKLVLLRRETRNLVANKL 384
 Db 368 MHRSEEVKQTKRLKGL-----RKDSSMEGCHSFASOLJTGFCILFKFTPLSIMDSV 421
 QY 385 AVITRLQNLIMGLFLLFPLVLRVRSNVLKCAIDRGLLQENGAIPYTGMLNANLVP 444
 Db 422 LTHLRITSHIGLIGLLGLVIGLIGNEAK--VLNSGFFLPFSLMFLMFLALMPTVLTFPL 479
 QY 445 LRAVSDQSDGLYQKQMLAYALHVLPPSVATMFFSSVCYWTGLHGEVARFCFSA 504
 Db 480 EMGVFLREHLNMYSLKAYTLAKTMADVPOIMEPVAVCSIVYMTSOPSDAVRFVLEFA 539
 QY 505 ALAPHLIGFELVLVLLGIYQNPVNSVALLSINGVLVSGFLRNIOEMPIPKIISY 564
 Db 540 LGTMTSLVAOSLGL-LIGAASSTLQVATEVGPVTAIPVLLFSGFVSPDIPIYLOMXY 598
 QY 565 FTFQKCYSELVYNEPGLNFTGSSNVSTTTPMCAFTGQIOFIETCPGATSRPTMF 624
 Db 599 ISYVRIGFEGVILS-ITGLD---REDLHCDIDETCHF-OKSEAILRELDVENAKLYLDF 652
 QY 625 LILYSEFALVIGIYV--FKIR 645
 Db 653 IVLGIFPISRLAIYFVLRKIR 675
 QY
 Db
 RESULT 14
 WHIT_CERCA
 ID WHIT_CERCA STANDARD: PRT: 679 AA.
 AC 017320:
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE White protein.
 GN W.
 OS Ceratitis capitata (Mediterranean fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Tephritidae; Tephritidae; Ceratitis.
 OC NCBI_TaxId=7213;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-96123276; PubMed-8533095;
 RA Zwiabel L.J., Saccone G., Zacharopoulou A., Besansky N.J.,
 RA Favia G., Collins F.H., Louis C., Kafatos F.C.;
 RT The white gene of Ceratitis capitata: a phenotypic marker for
 RT germine transformation *;
 RT Science 270:2005-2007(1995).
 RL
 CC -I- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
 CC NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
 CC CELLS RESPONSIBLE FOR EYE COLOR.
 CC -I- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -I- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X89933; CAA61998.1;
 CC InterPro: IPR003439; ABC_transportr.
 CC InterPro: IPR005284; Pigment_permease.
 CC Pfam: PF00005; ABC_tran; 1.
 CC ProDom: PD000006; ABC_transportr; 1.
 CC TrIRFams: TrIRF00955; 3a01204; 1.

DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KW Pigment; ATP-binding; Transmembrane; Transport.
 FT NP_BIND 121 128 ATP (BY SIMILARITY).
 FT TRANSMEM 427 445 POTENTIAL.
 FT TRANSMEM 457 477 POTENTIAL.
 FT TRANSMEM 507 525 POTENTIAL.
 FT TRANSMEM 534 555 POTENTIAL.
 FT TRANSMEM 568 586 POTENTIAL.
 FT TRANSMEM 651 670 POTENTIAL.
 FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SEQUENCE 679 AA; 75145 MW; 3F9CBCTBA835C4OC CRC4;
 Query Match 17.8%; Score 591; DB 1; Length 679;
 Best Local Similarity 28.4%; Pred. No. 9,4e-34;
 Matches 176; Conservative 125; Mismatches 231; Indels 88; Gaps 18;
 QY 66 ROLIKDVSLVESGOINICLIGSSGCTTLLDMSGRIGRAGTFLEGV--YVNGRLRRE 123
 Db 101 KHLKNDSCVAYFGELLAVGSSGAGKTLTLNLSAFRSSGVOISPTIRMLGHPVDAR 160
 QY 124 QFQDCESYVLOSDTLSSLVRETLHYTALLAIRGNPSPFOK--KVEANMAELSHVA 181
 Db 161 EMQARCAVYQODDLFGSLTARHLLFQANVRMR-HMQOKQVQVQVDOVLODSLCKGO 219
 QY 182 DRIG-NYSLGGISTEERRRVSIAAQLDDPKVMLFDEPTGLDCKTANOIVVLELAR 240
 Db 220 NTLIGVGRYKGGISGEERKRLAFASBALDPPILLIDEPGLDSEMAHSVQVLEKLSQ 279
 QY 241 RNRIVLTIHOPSELEFQLEDKTALISFCELLFCGPAEMDFENCCGTPCEHNSPPOF 300
 Db 280 KGTIVLTIHOPSELEFELDKTLMAEGVAFGLTGAADVFEYTGATCPNTYPAOF 339
 QY 301 YNDLSVDTQSKEREIETSKRYOMIESAYKSAIKHTLKNIERMKHLTLPVPEKTD 360
 Db 340 YVOVLAV---VPQREVESDRVAKIDNRAVGVSKEMQNGQ----KLVSNGCKED 391
 QY 361 -----SPGVSKLVLLRVRNVLNKLAVITRLQNLIMGLFL-LEFVLVRVSNV 413
 Db 392 ENERTYKASWFMQFRAVLWRMSLVLEKELVAVRLQTTMAVILGLIFLCOQLTV-- 449
 QY 414 GAIODRVLGLYFOVGATPYTGMLNANLFPVLRAVSDOSOGIYQKQMLAYALHVP 473
 Db 450 -GYMANINGAIFLELTMTQNSPATYVFTTELPMFMRRLRRCDTYFFLGKTYIAELP 508
 QY 474 FSVVATMFFSSVCYWTGLHPEVARGYSAALLAPHLIGFELVLLGIYQNPVNSV 533
 Db 509 LFLVVPFLFTALAYPLIGLRPVDH-----FTALALVTLV--ANVSTSF 551
 QY 534 VALLS-----IAGVLVSGFLRNIOEMDIPKIIISYFFQKCYSELV 577
 Db 552 GYLISQACSSSTMAHSVGPVYIPELLEGGFELNLSGVYVYFKMLSYLSMFRANGLLI 611
 QY 578 NEFTGL--NFTGSSNVSTTTPMCAFTGQIOFIETKCP--GANSRFTMNL---ILKS 629
 Db 612 NQWADYKPELTCTLSNT-----TCPSGGEVILETLNFSASDLPD 652
 QY 630 FIP-ALVITIGIVFKIRDL 648
 Db 653 FIGLALLIG---PRISAVI 669
 QY
 Db
 RESULT 15
 ABG4_HUMAN
 ID ABG4_HUMAN STANDARD: PRT: 646 AA.
 AC 09H172:
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 4.
 GN ABCG4 OR WHITR2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia: Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RX SEQUENCE FROM N.A.
 RA MEDLINE-2518231; PubMed-1160608;
 RA Engel T., Lorkowaki S., Lueken A., Rust S., Schlucter B., Berger G.,
 RA Cullen P., Asseman G.;
 RT "The human ABCG4 gene is regulated by oxysterols and retinoids in
 RT monocyte-derived macrophages";
 RL Biochem. Biophys. Res. Commun. 288:483-488(2001).
 RN [2]
 RP SEQUENCE OF 20-646 FROM N.A.
 RC TISSUE-dorsal root ganglion;
 RA Oldfield S., Lowry C.A., Lightman S.L.;
 RT "Cloning and expression of a novel mammalian white family
 RT ABC-transporter: WHITE2";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: May be involved in macrophage lipid homeostasis.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AJ300465; CAC17140.1; -.
 DR Genbank: HGNC:13884; ABCG4.
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_Transporter.
 DR Pfam: PF00005; ABC_tran: 2.
 DR ProDom: PD000006; ABC_transportr: 1.
 DR SMART: SM00382; AAA: 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER: 1.
 DR ATP-binding; Glycoprotein; Transmembrane; Transport.
 FT DOMAIN 1 393
 FT TRANSMEM 394 414
 FT DOMAIN 415 425
 FT TRANSMEM 426 446
 FT DOMAIN 447 472
 FT TRANSMEM 473 493
 FT DOMAIN 494 503
 FT TRANSMEM 504 524
 FT DOMAIN 525 532
 FT TRANSMEM 533 553
 FT DOMAIN 554 617
 FT TRANSMEM 618 638
 FT DOMAIN 639 646
 FT NP_BIND 102 109
 FT CARBOHYD 422 422
 SQ SEQUENCE 646 AA: 71895 MW: 90CECE6E150772611 CRC64;
 Query Match 17.4%; Score 578.5; DB 1; Length 646;
 Best Local Similarity 27.1%; Pred. No. 6.6e-33;
 Matches 171; Conservative 126; Mismatches 274; Indels 59; Gaps 14;
 Oy 33 PEPSSGLHMSYSVSHRVRPMDITSCROQMTQIILKDVSLYVSGOIMCIGSSGSGK 92
 Db 54 PKRSADVLEFVELSYVREPCW-----RRKGYKTLKCLSGKFCRRELIGIMGPGSGK 108
 Oy 93 TTLDAASGRLAGTGEYVYNGRALRREPODFCSYVQSDTLLSSLVRETLHYTA 152
 Db 109 STEFNILAGY--RESGKGGILLVNGRPRELTFKMSCITIQDDMLPPLTLVLEAMVSA 166
 Oy 153 LLAIRGNPGSFGKKEAV-----MAELSLSHVADRILGNYSLGISTGERRRVSIAAQ 206
 Db 167 NLKLSERK---EVKKELVETILTALGIMSCSHRTAL-----LSGGGRKRLATALE 214
 Oy 207 LLDPPKVMLEDEPTTGDCMTANQIVLVLELARRNKIVLTITQPSSELFOLDKTAAIL 266

Db 215 LVNPPVAFDEPTSGLDASCFQVVSILMKSIAOGRTIITIHQPSAKLEFEMDKLTL 274
 Oy 267 SEGELIFGCTPAEMDEFNDGPGCPHEHSNPFDDYMDLSTVDYOSKEEITSKRVOMIE 326
 Db 275 SOGQCTFGVVTNLLPYLKGJGLHCPTIHNPADETLIEVASC-----EYGLNPLFL 325
 Oy 327 SAYKSAICHKHTLKNIEPRKHLKTLPMVPFETKSP-----GVFSKLGVLRRVTRN 378
 Db 326 RAV-QNGICAAAEKSSPERKNEVAPCPCPPEVDPIESHFASTLTQPCILKRTFLS 384
 Oy 379 LVNKLAVITRLQNLINGELFLFEVLRVRSNVLKGAIQDRVGLYQFVGATPYTGMLNA 438
 Db 385 ILRDVLTHLRFMSHVIVGLIGLTYLHIGDASK--VENNTGCLFSSMLFLMRALMPT 442
 Oy 439 VNLPPVLAASVDQSDGLYOKMOMLALVLPFSVATMISSCYTWTGLHPEYAR 498
 Db 443 VLTPPLEAVPMREHLNTWYSILKAYYLAKTADVDPEQVVCVVCSTIYVNTGCPAETS 502
 Oy 499 FGYSALLAHLIGEFLTLVLIVGIONPNIYNSVVALSLAGVYVSGFLRNTQEMP 558
 Db 503 FLIFSALATATALVAQSLGL-LIGASNSLOVATFVGVTALPVLLSGFVSFKTTPY 561
 Oy 559 FKIIISYFPQYCEELVYVNEFYGL--NFTCGSSNVSVTNPMCAFTQGIQFTEKCPG 615
 Db 562 LQMSYLSYVRYGFEVILT-IYGMEREDLTC-----LEERCPPREP-QSILRALDV 611
 Oy 616 ATRSMTEFLILYSITPALTGLIVYKIR 645
 Db 612 EDARKYMDFLVIGIFLALRLALVLAIVLYR 641

Search completed: June 11, 2003, 09:12:37
 Job time : 18 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 09:04:17 ; Search time 41 Seconds

(Without alignments)
2115.760 Million cell updates/sec

Title: US-09-989-981a-6

Sequence: 1 MCDLSILTRGSGMGLVNRG.....PALVIIGIVFKIRDLISR 651

Scoring table: BLOSUM62

Gap 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: A: Geneseq_101002.*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3336	100%	651	23	Human ABCG5 protei
2	3326	100%	651	23	Human ABCG5 protei
3	3323	99.9	651	23	Human ABCG5 protei
4	3321	99.8	651	23	Human ABCG5 protei
5	3321	99.8	651	23	Human ABCG5 protei
6	3319	99.8	651	23	Human ABCG5 protei
7	2748.5	82.6	652	23	Human ABCG5 protei
8	2744.5	82.5	652	23	Human ABCG5 protei
9	2742.5	82.5	652	23	Human ABCG5 protei
10	2738.5	82.3	652	23	Human ABCG5 protei

11	2727.5	82.0	652	23	Human ABCG5 protei
12	2081	62.6	408	23	Human ABCG5 protei
13	1387.5	41.7	340	23	Human ABCG5 protei
14	1234	37.1	243	23	Human ABCG5 protei
15	722	21.7	144	23	Human ABCG5 protei
16	682.5	20.5	655	23	Human ABCG5 protei
17	682.5	20.5	655	23	Human ABCG5 protei
18	680.5	20.5	655	23	Human ABCG5 protei
19	680.5	20.5	655	23	Human ABCG5 protei
20	680.5	20.5	655	23	Human ABCG5 protei
21	680.5	20.5	655	23	Human ABCG5 protei
22	676.5	20.3	665	23	Human ABCG5 protei
23	674.5	20.3	655	23	Human ABCG5 protei
24	674.5	20.3	655	23	Human ABCG5 protei
25	674.5	20.3	655	23	Human ABCG5 protei
26	672.5	20.2	655	23	Human ABCG5 protei
27	660	19.8	657	23	Human ABCG5 protei
28	602.5	18.1	602	22	Human ABCG5 protei
29	600	18.0	602	22	Human ABCG5 protei
30	588.5	17.7	648	21	Human ABCG5 protei
31	587.5	17.7	635	21	Human ABCG5 protei
32	587.5	17.7	632	21	Human ABCG5 protei
33	581.5	17.5	687	21	Human ABCG5 protei
34	578.5	17.4	646	23	Human ABCG5 protei
35	578	17.4	604	20	Human ABCG5 protei
36	578	17.4	604	20	Human ABCG5 protei
37	575.5	17.3	649	21	Human ABCG5 protei
38	575.5	17.3	653	21	Human ABCG5 protei
39	571	17.2	609	22	Human ABCG5 protei
40	569.5	17.1	666	23	Human ABCG5 protei
41	564.5	17.0	616	21	Human ABCG5 protei
42	553	16.6	666	22	Human ABCG5 protei
43	515.5	15.5	832	22	Human ABCG5 protei
44	473	14.2	689	22	Human ABCG5 protei
45	463.5	13.9	812	22	Human ABCG5 protei

ALIGNMENTS

RESULT 1	AAU96984	standard; Protein; 651 AA.
ID	AAU96984	
AC	AAU96984	
DT	30-JUL-2002	(first entry)
XX		
DE	Human ABCG5 protein.	
XX		
KW	Human: ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol; arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease; chromosome 2p21.	
KW		
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	Misc-difference 2..15	/note="Encoded by GRCYC"
XX		
PD	04-APR-2002.	
XX		
PF	25-SEP-2001; 2001MO-US29859.	
XX		
PR	25-SEP-2000; 2000US-235268P.	
XX		
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.	
PA	(PATE/) PATEL S B.	
PA	(DEAN/) DEAN M.	
XX		
PI	Patel SB, Dean M.	

XX WPI; 2002-416483/44.
 DR N-PSDB; ABR51681.
 XX Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 XX sitosterolemia, arteriosclerosis and heart diseases
 PS Claim 52; Page 35-36; 66pp; English.

CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 protein of the invention. This
 CC sequence is encoded by the human ABCG5 gene located on chromosome 2p21.
 XX
 XX
 SQ Sequence 651 AA;

Query Match 100.0%; Score 3326; DB 23; Length 651;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MGDLSTLPGGSMGLQVNRGSSSLGAPATAPAPHSGLTHASVSHRRVPMDDISC 60
 DB 1 MGDLSTLPGGSMGLQVNRGSSSLGAPATAPAPHSGLTHASVSHRRVPMDDISC 60
 OY 61 RQOWTROLKLVSLYVESGQIMCITIGSSGKTTLLDMSGRILGAGTFLEVVYNGRAL 120
 DB 61 RQOWTROLKLVSLYVESGQIMCITIGSSGKTTLLDMSGRILGAGTFLEVVYNGRAL 120
 OY 121 RREQFODCFYVLSGDTLLSLVYRETHYATALLAIRGNNGSPQKTEAVMAELTSHV 180
 DB 121 RREQFODCFYVLSGDTLLSLVYRETHYATALLAIRGNNGSPQKTEAVMAELTSHV 180
 OY 121 RREQFODCFYVLSGDTLLSLVYRETHYATALLAIRGNNGSPQKTEAVMAELTSHV 180
 DB 121 RREQFODCFYVLSGDTLLSLVYRETHYATALLAIRGNNGSPQKTEAVMAELTSHV 180
 OY 181 ADRLIGNYSLGISTGERRRVSIQAQLQDPKVMLEDEPTGLDQMTANQIVLLVELAR 240
 DB 181 ADRLIGNYSLGISTGERRRVSIQAQLQDPKVMLEDEPTGLDQMTANQIVLLVELAR 240
 OY 241 RNRIVVLTHIRPRESLRDLFKIALISGELIFCGTFAEMLDFNDCGYPCEHSNPDE 300
 DB 241 RNRIVVLTHIRPRESLRDLFKIALISGELIFCGTFAEMLDFNDCGYPCEHSNPDE 300
 OY 241 RNRIVVLTHIRPRESLRDLFKIALISGELIFCGTFAEMLDFNDCGYPCEHSNPDE 300
 DB 241 RNRIVVLTHIRPRESLRDLFKIALISGELIFCGTFAEMLDFNDCGYPCEHSNPDE 300
 OY 301 YMDLTSVDTSKEREIETSKRWOMIESAYKKSATCKLKNIERKHKTLPMPPEKTKD 360
 DB 301 YMDLTSVDTSKEREIETSKRWOMIESAYKKSATCKLKNIERKHKTLPMPPEKTKD 360
 OY 301 YMDLTSVDTSKEREIETSKRWOMIESAYKKSATCKLKNIERKHKTLPMPPEKTKD 360
 DB 301 YMDLTSVDTSKEREIETSKRWOMIESAYKKSATCKLKNIERKHKTLPMPPEKTKD 360
 OY 361 SPGVFSKILGVLRLRRVRLVNRKLAIVTRLQNLIMGLFLEFVLAVRSNVLKAIODRV 420
 DB 361 SPGVFSKILGVLRLRRVRLVNRKLAIVTRLQNLIMGLFLEFVLAVRSNVLKAIODRV 420
 OY 421 GLTYQFGATPYTGMNAVNLFPLVRAVSDQESDGLYQKQMMALAYALHVLPSVATM 480
 DB 421 GLTYQFGATPYTGMNAVNLFPLVRAVSDQESDGLYQKQMMALAYALHVLPSVATM 480
 OY 481 IFSSVCWTGLGHEVVARFGYFSAALLAPHIGSEFVLVLTGYONRNIVASVALLSIA 540
 DB 481 IFSSVCWTGLGHEVVARFGYFSAALLAPHIGSEFVLVLTGYONRNIVASVALLSIA 540

OY 541 GVLVSGFRLNIOEMPIPEKTIISYTPFKYCSLIVWVEFGNLNCCSSNVSYTNPMC 600
 DB 541 GVLVSGFRLNIOEMPIPEKTIISYTPFKYCSLIVWVEFGNLNCCSSNVSYTNPMC 600
 OY 601 AFTGCIQPIETKPCGATSRFTWNLILYSFIPALVILIVFKIRIDHLISR 651
 DB 601 AFTGCIQPIETKPCGATSRFTWNLILYSFIPALVILIVFKIRIDHLISR 651

RESULT 2
 AAE13290
 ID AAE13290 standard; Protein; 651 AA.
 AC AAE13290;
 DE 12-FEB-2002 (first entry)

Human sitosterolemia susceptibility gene (SSG) protein.

Human; sitosterolemia susceptibility gene; SSG; arteriosclerosis;
 KW sterol-related disorder; hyperlipidaemia; hypercholesterolemia; therapy;
 KW gall stone; coronary heart disease; cardiovascular disease; arthritis;
 KW xanthoma; haemolytic anaemia; transgenic animal; chromosome 2p21.

OS Homo sapiens.

W0200179272-42.
 25-OCT-2001.

18-APR-2001; 2001WO-US12758.

18-APR-2000; 2000US-198465P.

15-MAY-2000; 2000US-204234P.

(TULIA-) TULARIK INC.

Tian H, Schultz J, Shan B;

WPI; 2002-017598/02.

N-PSDB; AAD22009.

Novel sitosterolemia susceptibility gene polypeptide and
 PT polynucleotide, useful for screening a compound that increases the
 PT level of expression or activity of SSG polypeptide for treating
 PT sterol-related disorder

Claim 19; Fig 8; 105pp; English.

The invention relates to an isolated sitosterolemia Susceptibility Gene
 (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
 binding cassette (ABC) family cholesterol transporter. SSG is useful
 CC for identifying a compound useful in the treatment or prevention of a
 CC sterol-related disorder, including sitosterolemia, hyperlipidaemia,
 CC hypercholesterolemia, gall stones, HDL deficiency, arteriosclerosis or
 CC nutritional deficiencies. SSG is also useful for treating cholesterol-
 CC associated diseases or conditions including coronary heart disease and
 CC other cardiovascular diseases, and sitosterolemia-associated condition
 CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
 CC expression cassette is useful in the production of transgenic non-human
 CC animals. SSG genes and their homologues are useful as tools for a number
 CC of applications including diagnosing sitosterolemia and other
 CC cardiovascular disorders, for forensics and paternity determinations,
 CC and for treating any of a large number of SSG associated diseases. The
 CC present sequence is human SSG protein. Human SSG is located on chromosome
 CC 2p21.

Sequence 651 AA;

Query Match 100.0%; Score 3326; DB 23; Length 651;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 MGDLSLTPGSGMGLQVNRGSSSLGEGAPATAPAPPHSGILHASTSVSHRVRPMDITSC 60
DB 1 MGDLSLTPGSGMGLQVNRGSSSLGEGAPATAPAPPHSGILHASTSVSHRVRPMDITSC 60
OY 61 RQWRTQILKDVSLYVESGOIMCITLGGSSGKTTLLDMSGRGLRAGFTLGEVYNGAL 120
DB 61 RQWRTQILKDVSLYVESGOIMCITLGGSSGKTTLLDMSGRGLRAGFTLGEVYNGAL 120
OY 121 RREPFOCFSTVLDSTLSSLYRETLHTATLAIIRGNBSPCKVEAVMAELSLSHV 180
DB 121 RREPFOCFSTVLDSTLSSLYRETLHTATLAIIRGNBSPCKVEAVMAELSLSHV 180
OY 181 ADRLIGNSLGISTGERRRVSIQAQLLQDPKVMLEDEPTTGGLDQMTANOIVLVLELAR 240
DB 181 ADRLIGNSLGISTGERRRVSIQAQLLQDPKVMLEDEPTTGGLDQMTANOIVLVLELAR 240
OY 241 RNRIVVLTIHQPRSELQPLDKIALISFGEILFCGTPAEMLDGFNDGCPCEHSNPDF 300
DB 241 RNRIVVLTIHQPRSELQPLDKIALISFGEILFCGTPAEMLDGFNDGCPCEHSNPDF 300
OY 301 YMDLTSVDTSKEREIETSKRVOMIESAYKKSALCHKTLKNIERKHLKTLPMVPFKTKD 360
DB 301 YMDLTSVDTSKEREIETSKRVOMIESAYKKSALCHKTLKNIERKHLKTLPMVPFKTKD 360
OY 361 SPGVSKLGVLLRVTNRLVNRKLAIVITRLQNLIMGLFLFVLRVRSNVLKGAIDRV 420
DB 361 SPGVSKLGVLLRVTNRLVNRKLAIVITRLQNLIMGLFLFVLRVRSNVLKGAIDRV 420
OY 421 GLTQFGATPYTGGLNAVNLPVLRVAVSDSDGIXOKOMMLAYALHVPFSVATM 480
DB 421 GLTQFGATPYTGGLNAVNLPVLRVAVSDSDGIXOKOMMLAYALHVPFSVATM 480
OY 481 IFSSVCTWTGLHPEVNAFGYFSAAALAPHILGELTLLVGLIYQNPINVSVALSLIA 540
DB 481 IFSSVCTWTGLHPEVNAFGYFSAAALAPHILGELTLLVGLIYQNPINVSVALSLIA 540
OY 541 GVLVSGFLRNIOEMPIPKIISYTFQKCYSELIYVNEFGNLFTGSSNVSATYTNMC 600
DB 541 GVLVSGFLRNIOEMPIPKIISYTFQKCYSELIYVNEFGNLFTGSSNVSATYTNMC 600
OY 601 AFTOGIOPIEKTCPGATSRFTMNFLLIYSFIPALVILGIVFKIRDLHISR 651
DB 601 AFTOGIOPIEKTCPGATSRFTMNFLLIYSFIPALVILGIVFKIRDLHISR 651

```

RESULT 3
AAU96992
ID AAU96992 standard; Protein: 651 AA.
AC AAU96992;
XX
DT 30-JUL-2002 (first entry)
XX
DE Human ABCG5 mutant E146Q protein sequence.
XX
KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
KW mutant; muten.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key location/Qualifiers
FT Misc-difference 146 /note- "wild-type Glu substituted by Gln"
XX
PD 04-APR-2002.
PF 25-SEP-2001; 2001WO-0529859.
XX

```

PR 25-SEP-2000; 2000US-235268P.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA (PATR/) PATEL S B.
PA (DEAN/) DEAN M.
XX
XX Patel SB, Dean M;
XX
XX WPI: 2002-416483/44.
XX
XX Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
PT nucleic acid encoding the polypeptide, useful for treating
PT sitosterolemia, arteriosclerosis and heart diseases
XX
PS Claim 12; Page -- 66pp; English.
XX
XX The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying
CC a compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present amino
CC acid sequence represents the human ABCG5 mutant E146Q protein of the
CC invention.
CC Note: This sequence is not shown in the specification but is derived
CC from the wild-type human ABCG5 protein (AAU96984) given on pages 35-36
CC of the specification.
XX
XX Sequence 651 AA:
XX
XX Query Match 99.9%; Score 3323; DB 23; Length 651;
XX Best Local Similarity 99.8%; Pred. No. 0;
XX Matches 650; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

1 MGDLSLTPGSGMGLQVNRGSSSLGEGAPATAPAPPHSGILHASTSVSHRVRPMDITSC 60
1 MGDLSLTPGSGMGLQVNRGSSSLGEGAPATAPAPPHSGILHASTSVSHRVRPMDITSC 60
61 RQWRTQILKDVSLYVESGOIMCITLGGSSGKTTLLDMSGRGLRAGFTLGEVYNGAL 120
61 RQWRTQILKDVSLYVESGOIMCITLGGSSGKTTLLDMSGRGLRAGFTLGEVYNGAL 120
121 RREPFOCFSTVLDSTLSSLYRETLHTATLAIIRGNBSPCKVEAVMAELSLSHV 180
121 RREPFOCFSTVLDSTLSSLYRETLHTATLAIIRGNBSPCKVEAVMAELSLSHV 180
181 ADRLIGNSLGISTGERRRVSIQAQLLQDPKVMLEDEPTTGGLDQMTANOIVLVLELAR 240
181 ADRLIGNSLGISTGERRRVSIQAQLLQDPKVMLEDEPTTGGLDQMTANOIVLVLELAR 240
241 RNRIVVLTIHQPRSELQPLDKIALISFGEILFCGTPAEMLDGFNDGCPCEHSNPDF 300
241 RNRIVVLTIHQPRSELQPLDKIALISFGEILFCGTPAEMLDGFNDGCPCEHSNPDF 300
301 YMDLTSVDTSKEREIETSKRVOMIESAYKKSALCHKTLKNIERKHLKTLPMVPFKTKD 360
301 YMDLTSVDTSKEREIETSKRVOMIESAYKKSALCHKTLKNIERKHLKTLPMVPFKTKD 360
361 SPGVSKLGVLLRVTNRLVNRKLAIVITRLQNLIMGLFLFVLRVRSNVLKGAIDRV 420

Db 361 SPGVSKLGLVLRVTRNLRNKLAVITRLLQNLIMGLFLFFVLRVSNVKGAIQDRV 420
 Qy 421 GLLYQFVGATPYTGMLNANVLPVLRVAVSDQSDGLYOKWOMLAVALHVLPEFSVATM 480
 Db 421 GLLYQFVGATPYTGMLNANVLPVLRVAVSDQSDGLYOKWOMLAVALHVLPEFSVATM 480
 Qy 481 IFSSVCYWTGLGHPVAVRGYFSAALAPHLIGEFLLVLLGIVONPNVNSVALLSTA 540
 Db 481 IFSSVCYWTGLGHPVAVRGYFSAALAPHLIGEFLLVLLGIVONPNVNSVALLSTA 540
 Qy 541 GVLVSGFLRNIOEMPPIKIIISYFFQKCYSEILVNEFYGLNFTCGSSNSVYTNPMPC 600
 Db 541 GVLVSGFLRNIOEMPPIKIIISYFFQKCYSEILVNEFYGLNFTCGSSNSVYTNPMPC 600
 Qy 601 AFTQGIQFIKTCPCATSRFTMNFLLYSFIPALVILGIVVKIRIDHLISR 651
 Db 601 AFTQGIQFIKTCPCATSRFTMNFLLYSFIPALVILGIVVKIRIDHLISR 651

RESULT 4
 AA096989 standard; Protein; 651 AA.
 AC AA096989;
 XX
 DT 30-JUL-2002 (first entry)
 DE Human ABCG5 mutant R419H protein sequence.
 XX
 KM Human ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol; arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease; mutant; mutin.
 OS Homo sapiens.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 419 /note="Wild-type Arg substituted by His"
 FT
 PN MO200227016-A2.
 PD 04-APR-2002.
 XX
 PF 25-SEP-2001; 2001WO-US29859.
 XX
 PR 25-SEP-2000; 2000US-235268P.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 XX
 PI Patel SB, Dean M;
 XX
 DR MPI; 2002-416483/44.
 XX
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 XX
 PS Claim 9: Page -: 66pp: English.

The present invention relates to a new mammalian ATP-binding cassette gene 5 (ABCG5) polypeptide. The invention is useful for identifying a predisposition for developing sitosterolemia, arteriosclerosis or heart disease. The molecules of the invention are also useful for identifying a compound which alters ABCG5 activity level comprising contacting a cell culture or mammal which have ABCG5 polypeptide with a compound and measuring ABCG5 biological activity in the cell culture or in mammal, where an increase or decrease in ABCG5 biological activity compared to CC ABCG5 biological activity in a control cell culture or mammal not contacted with the compound, identifies a compound that increases or decreases ABCG5 activity respectively. The cell culture or mammal

CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 mutant R419H protein of the
 CC invention.
 CC Note: This sequence is not shown in the specification but is derived
 CC from the wild-type human ABCG5 protein (AA096984) given on pages 35-36
 CC of the specification.
 XX
 SQ Sequence 651 AA;
 Query Match 99.88; Score 3321; DB 23; Length 651;
 Best Local Similarity 99.88; Pred. No. 0;
 Matches 650; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 MGDLSLTPGSGMLQVNRGSSSLGAPATAPRPHSLGILHNSYSVSHVRPMDITSC 60
 Db 1 MGDLSLTPGSGMLQVNRGSSSLGAPATAPRPHSLGILHNSYSVSHVRPMDITSC 60
 Qy 61 RQOMTRQILKDVSLVYESGQIMCILLSSGSGKTTLLDAMSGRLGRAGTFAGEYVNGRAL 120
 Db 61 RQOMTRQILKDVSLVYESGQIMCILLSSGSGKTTLLDAMSGRLGRAGTFAGEYVNGRAL 120
 Qy 121 RREQFQDCSTYVLOSPTLLSLTVRETLHTALAIRGNPNSFORKEVAVMAELSLSHV 180
 Db 121 RREQFQDCSTYVLOSPTLLSLTVRETLHTALAIRGNPNSFORKEVAVMAELSLSHV 180
 Qy 181 ADRLIGNYSLGSGSTERRRVSIAAQLDPPKMLFDEPTTGIDCTANQIVLVLELAR 240
 Db 181 ADRLIGNYSLGSGSTERRRVSIAAQLDPPKMLFDEPTTGIDCTANQIVLVLELAR 240
 Qy 241 RNRIVVLTIHQPSSELFQLEDKAIILSFGLIRCGTAPAEMLDFENCGYPCPEHNPDE 300
 Db 241 RNRIVVLTIHQPSSELFQLEDKAIILSFGLIRCGTAPAEMLDFENCGYPCPEHNPDE 300
 Qy 301 YMDLISVDTQSKEREIETSKRVQMIESAVKRSACHRTLKNIERMKHLTLPVAVPKTMD 360
 Db 301 YMDLISVDTQSKEREIETSKRVQMIESAVKRSACHRTLKNIERMKHLTLPVAVPKTMD 360
 Qy 361 SPGVSKLGLVLRVTRNLRNKLAVITRLLQNLIMGLFLFFVLRVSNVKGAIQDRV 420
 Db 361 SPGVSKLGLVLRVTRNLRNKLAVITRLLQNLIMGLFLFFVLRVSNVKGAIQDRV 420
 Qy 421 GLLYQFVGATPYTGMLNANVLPVLRVAVSDQSDGLYOKWOMLAVALHVLPEFSVATM 480
 Db 421 GLLYQFVGATPYTGMLNANVLPVLRVAVSDQSDGLYOKWOMLAVALHVLPEFSVATM 480
 Qy 481 IFSSVCYWTGLGHPVAVRGYFSAALAPHLIGEFLLVLLGIVONPNVNSVALLSTA 540
 Db 481 IFSSVCYWTGLGHPVAVRGYFSAALAPHLIGEFLLVLLGIVONPNVNSVALLSTA 540
 Qy 541 GVLVSGFLRNIOEMPPIKIIISYFFQKCYSEILVNEFYGLNFTCGSSNSVYTNPMPC 600
 Db 541 GVLVSGFLRNIOEMPPIKIIISYFFQKCYSEILVNEFYGLNFTCGSSNSVYTNPMPC 600
 Qy 601 AFTQGIQFIKTCPCATSRFTMNFLLYSFIPALVILGIVVKIRIDHLISR 651
 Db 601 AFTQGIQFIKTCPCATSRFTMNFLLYSFIPALVILGIVVKIRIDHLISR 651

RESULT 5
 AA096990 standard; Protein; 651 AA.
 AC AA096990;
 XX
 DT 30-JUL-2002 (first entry)
 XX

DE Human ABCG5 mutant R389H protein sequence.

XX Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KM mutant; mutlein.

XX Homo sapiens.
 OS Synthetic.

XX Key Location/Qualifiers
 FT Misc-difference 389 /note- "Wild-type Arg substituted by His"
 FT FT
 XX W0200227016-A2.
 XX 04-APR-2002.
 XX 25-SEP-2001; 2001WO-US29859.
 XX 25-SEP-2000; 2000US-235268P.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 XX Patel SB, Dean M;
 PI WPI; 2002-416483/44.
 DR Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 XX
 PS Claim 7; Page -: 66pp; English.

CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 mutant R389H protein of the
 CC invention.
 CC Note: This sequence is not shown in the specification but is derived
 CC from the wild-type human ABCG5 protein (AA096984) given on pages 35-36
 CC of the specification.
 XX
 SQ Sequence 651 AA;

Query Match 99.8%; Score 3321; DB 23; Length 651;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 650; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGLDLSITPGSGMGLVNRGSSSLGAPATAPENHSLGILHASYSHRYRPMWDTSC 60
 DB 1 MGLDLSITPGSGMGLVNRGSSSLGAPATAPENHSLGILHASYSHRYRPMWDTSC 60
 QY 61 RQGWTRQILKDVSLVYSSQIMKILGSSGSKTTLIDAMSGRLGRAGTFAGYVNGRAL 120
 DB 61 RQGWTRQILKDVSLVYSSQIMKILGSSGSKTTLIDAMSGRLGRAGTFAGYVNGRAL 120

QY 121 RREPOFQDESVYLOSDTLSSLTRETIHTYATALLAIRGNPGSPQKKEAYVAEISLSHV 180
 DB 121 RREPOFQDESVYLOSDTLSSLTRETIHTYATALLAIRGNPGSPQKKEAYVAEISLSHV 180
 QY 181 ADRLIGNSYLGGISTGERRRVSIQAQLQDPKVMLEDPPTGLDPMTANQIVLVLELAR 240
 DB 181 ADRLIGNSYLGGISTGERRRVSIQAQLQDPKVMLEDPPTGLDPMTANQIVLVLELAR 240
 QY 241 RNRIVYLTIRHQRSELFDLFDKIALISGELIFCGTPAEMLDFNDCCYPCPEHSPNPDF 300
 DB 241 RNRIVYLTIRHQRSELFDLFDKIALISGELIFCGTPAEMLDFNDCCYPCPEHSPNPDF 300
 QY 301 YMDLTSVDTQSKERIEFTSKRVOMESAYKKSALCHKTLKNIERRKHLKTLPMVPFKTKD 360
 DB 301 YMDLTSVDTQSKERIEFTSKRVOMESAYKKSALCHKTLKNIERRKHLKTLPMVPFKTKD 360
 QY 361 SPGVFSKLGVLRLRYTRMLVNRKLAIVIRLQNLIMGLFLFVLRVRSNYLKGAIODRV 420
 DB 361 SPGVFSKLGVLRLRYTRMLVNRKLAIVIRLQNLIMGLFLFVLRVRSNYLKGAIODRV 420
 QY 421 GLTIQFVGATPYTGMLNANVLFVYLRAVSDQSDGLYQKQMMALAYALVLPSPVATM 480
 DB 421 GLTIQFVGATPYTGMLNANVLFVYLRAVSDQSDGLYQKQMMALAYALVLPSPVATM 480
 QY 481 IFSSVCWTGLHPEVNAFGEYSALLAPHLIGEFLLTVLGIYQNPINVSVALLSIA 540
 DB 481 IFSSVCWTGLHPEVNAFGEYSALLAPHLIGEFLLTVLGIYQNPINVSVALLSIA 540
 QY 541 GVLVSGFLRNIOEMPIPKIISYFTQKCYSELIYVNEFGNFTGSSNVSATYTPMC 600
 DB 541 GVLVSGFLRNIOEMPIPKIISYFTQKCYSELIYVNEFGNFTGSSNVSATYTPMC 600
 QY 601 AFTQIGQFIETKCPGATSRFTMNLILYSFIPALVILGIYVEKIRDLHLSR 651
 DB 601 AFTQIGQFIETKCPGATSRFTMNLILYSFIPALVILGIYVEKIRDLHLSR 651

RESULT 6
 ID AA096993 standard; Protein; 651 AA.
 AA096993;
 30-JUL-2002 (first entry)
 DE Human ABCG5 mutant R419P protein sequence.
 KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KM mutant; mutlein.
 XX Homo sapiens.
 OS Synthetic.

XX Key Location/Qualifiers
 FT Misc-difference 419 /note- "Wild-type Arg substituted by Pro"
 FT FT
 XX W0200227016-A2.
 XX 04-APR-2002.
 XX 25-SEP-2001; 2001WO-US29859.
 XX 25-SEP-2000; 2000US-235268P.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 XX Patel SB, Dean M;

DR WPI: 2002-416483/44.
 XX Novel mammalian ATP-binding cassette gene 5 polypeptide; and the
 PT nucleic acid encoding the polypeptide, useful for treating
 XX stitosterolemia, arteriosclerosis and heart diseases
 PS Claim 10; Page -: 66pp; English.

CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing stitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 mutant R419P protein of the
 CC invention.
 CC Note: This sequence is not shown in the specification but is derived
 CC from the wild-type human ABCG5 protein (A095984) given on pages 35-36
 CC of the specification.
 XX
 XX Sequence 651 AA:

Query Match 99.8%; Score 3319; DB 23; Length 651;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 650; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MODLSLTPFGSGMGLOVNRGSSSLEGAPATAPEPHSLGILLASYSVSRVPPMDITSC 60
 DB 1 MODLSLTPFGSGMGLOVNRGSSSLEGAPATAPEPHSLGILLASYSVSRVPPMDITSC 60
 QY 61 RQOMTROLIKDVSILVSEGOIMCIISSGSGKTTLLDMNSGLRAGTFLGEVYNGRAL 120
 DB 61 RQOMTROLIKDVSILVSEGOIMCIISSGSGKTTLLDMNSGLRAGTFLGEVYNGRAL 120
 QY 121 RREOPDCEVYVLOSDTLLSLTVEETLHTYLLAIRGNPSSFOKKEVAVAEISLSHY 180
 DB 121 RREOPDCEVYVLOSDTLLSLTVEETLHTYLLAIRGNPSSFOKKEVAVAEISLSHY 180
 QY 181 ADRLIGNTSIGISTGERRRVSAIAOQLODPVMLEFDEPTTGLDCTANQIYVLLVELAR 240
 DB 181 ADRLIGNTSIGISTGERRRVSAIAOQLODPVMLEFDEPTTGLDCTANQIYVLLVELAR 240
 QY 241 RNRIVVLTIHOPRSELPOLFDKIALISFCELIFCGTPAEMLEDFNDGCGPCPCHSNPDFE 300
 DB 241 RNRIVVLTIHOPRSELPOLFDKIALISFCELIFCGTPAEMLEDFNDGCGPCPCHSNPDFE 300
 QY 301 TMDLSVDPQSKEREIETSKRVOMIESAVKSAICHTKLTAKNERKHLKTLPMVPFKTKD 360
 DB 301 TMDLSVDPQSKEREIETSKRVOMIESAVKSAICHTKLTAKNERKHLKTLPMVPFKTKD 360
 QY 361 SPGVSKLGLVLRVTRNLVNRKLAIVITRLQNLIMGLFLFVLRVSNVYKGIADRV 420
 DB 361 SPGVSKLGLVLRVTRNLVNRKLAIVITRLQNLIMGLFLFVLRVSNVYKGIADRV 420
 QY 421 GLIYFVGATPTTGMANVNLFPVLRVAVSDQSGLYOKQWOMLAVLHLVFPFSVATM 480
 DB 421 GLIYFVGATPTTGMANVNLFPVLRVAVSDQSGLYOKQWOMLAVLHLVFPFSVATM 480
 QY 481 IFSSVCYTTLGLHPEVARGYSALLPLHIGELTIVLGIYONPNIVNSVVALLSIA 540
 DB 481 IFSSVCYTTLGLHPEVARGYSALLPLHIGELTIVLGIYONPNIVNSVVALLSIA 540

DB 481 IFSSVCYTTLGLHPEVARGYSALLPLHIGELTIVLGIYONPNIVNSVVALLSIA 540
 QY 541 GVLVSGFLRNIOEMPIPKIISYFFQKCEILVNVNEFGINFTGSSNSVYTPNPMC 600
 DB 541 GVLVSGFLRNIOEMPIPKIISYFFQKCEILVNVNEFGINFTGSSNSVYTPNPMC 600
 QY 601 AFTGIGIETKTCPGATSRFTNMFILYSFIPALVILGIVKIRIDHLISR 651
 DB 601 AFTGIGIETKTCPGATSRFTNMFILYSFIPALVILGIVKIRIDHLISR 651

RESULT 7
 ID AAE13309 standard; Protein; 652 AA.
 XX AAE13309;
 AC 12-FEB-2002 (first entry)
 DT
 DE Mouse stitosterolaemia susceptibility gene (SSG) protein variant #2.
 DE
 KW Mouse; stitosterolaemia susceptibility gene; SSG; atherosclerosis; mutant;
 KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; mutant;
 KW gall stone; coronary heart disease; cardiovascular disease; arthritis;
 KW xanthoma; haemolytic anaemia; transgenic animal; therapy; variant.
 XX
 OS Mus sp.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Misc-difference 28 /note= "Wild type Gly substituted with Ala"
 FT
 XX WO200179272-A2.
 XX
 PD 25-OCT-2001.
 PD 18-APR-2001; 2001WO-US12758.
 PF 18-APR-2001; 2000US-198465P.
 PR 15-MAY-2000; 2000US-204234P.
 XX
 PA (TULA-) TULARIK INC.
 PI Tian H, Schultz J, Shan B;
 PI
 DR WPI: 2002-017598/02.
 PT Novel stitosterolemia susceptibility gene polypeptide and
 PT polynucleotide, useful for screening a compound that increases the
 PT level of expression or activity of SSG polypeptide for treating
 PT sterol-related disorder
 PS Disclosure: Page -: 105pp; English.
 XX
 CC The invention relates to an isolated stitosterolaemia susceptibility gene
 CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
 CC binding cassette (ABC) family cholesterol transporter. SSG is useful
 CC for identifying a compound useful in the treatment or prevention of a
 CC sterol-related disorder, including stitosterolaemia, hyperlipidaemia,
 CC hypercholesterolemia, gall stones, HDL deficiency, atherosclerosis or
 CC nutritional deficiencies. SSG is also useful for treating cholesterol-
 CC associated diseases or conditions including coronary heart disease and
 CC other cardiovascular diseases, and stitosterolaemia-associated condition
 CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
 CC expression cassette is useful in the production of transgenic non-human
 CC animals. SSG genes and their homologues are useful as tools for a number
 CC of applications including diagnosing stitosterolaemia and other
 CC cardiovascular disorders, for forensics and paternity determinations,
 CC and for treating any of a large number of SSG associated diseases. The
 CC present sequence is mouse SSG protein variant obtained by replacing
 CC Gly28 with Ala.
 CC Note: The present sequence is not shown in the specification but is

OY 420 VGLLQVFGATPTTGLNANVNFVPLRAVSDSDGLYOKQMMLAYALHVPESVAT 479
 DB 421 VGLLQVFGATPTTGLNANVNFVPLRAVSDSDGLYOKQMMLAYALHVPESVAT 480
 OY 480 MIFSSVCYWTGLLHPEVAFGFSALLAPHLIGELFVLGIYONPNIVNSVALLSI 539
 DB 481 VIFSSVCYWTGLLHPEVAFGFSALLAPHLIGELFVLGIYONPNIVNSVALLSI 540
 OY 540 AGVLVSGFLRNIOEMPIFKIISYFTFOKCYSEILVNEFGYGLNFTCGSSNVSTYTNM 599
 DB 541 SGLLIGSGFIRNIOEMPIFKIISYFTFOKCYSEILVNEFGYGLNFTCGSSNVSTYTNM 600
 OY 600 CAPTGOIPIETKTCGATSRFTMNLILYSPALVILGIYVKIKIDHLISR 651
 DB 601 CAITGVQFIETKTCGATSRFTMNLILYSPALVILGIYVKIKIDHLISR 652

RESULT 9
 AAEL3308
 ID AAEL3308 standard; Protein: 652 AA.
 AC AAEL3308;
 DE 12-FEB-2002 (first entry)
 XX Mouse sitosterolemia susceptibility gene (SSG) protein variant #1.
 DE Mouse sitosterolemia susceptibility gene (SSG) protein variant #1.
 XX Mouse; sitosterolemia susceptibility gene; SSG; atherosclerosis; mutant;
 KM sterol-related disorder; hyperlipidaemia; hypercholesterolemia; mutant;
 KM gall stone; coronary heart disease; cardiovascular disease; arthritis;
 KM xanthoma; haemolytic anaemia; transgenic animal; therapy; variant.
 XX Mus sp.
 OS Synthetic.
 OS
 FH Key Location/Qualifiers
 FT Misc-difference 17
 FT /note= "Wild type Ile substituted with Leu"
 PN W0200179272-A2.
 PD 25-OCT-2001.
 PE 18-APR-2001; 2001WO-US12758.
 PR 18-APR-2000; 2000US-198465P.
 PR 15-MAY-2000; 2000US-204234P.
 XX (TULA-) TULARIK INC.
 PA
 PI Tian H, Schultz J, Shan B.
 DR WPI: 2002-017598/02.
 XX
 PT Novel sitosterolemia susceptibility gene polypeptide and
 PT polynucleotide, useful for screening a compound that increases the
 PT level of expression or activity of SSG polypeptide for treating
 PT sterol-related disorder
 XX
 PS Disclosure: Page -: 105pp; English.
 XX
 CC The invention relates to an isolated sitosterolemia Susceptibility Gene
 CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
 CC binding cassette (ABC) family cholesterol transporter. SSG is useful
 CC for identifying a compound useful in the treatment or prevention of a
 CC sterol-related disorder, including sitosterolemia, hyperlipidaemia,
 CC hypercholesterolemia, gall stones, HDL deficiency, atherosclerosis or
 CC nutritional deficiencies. SSG is also useful for treating cholesterol-
 CC associated diseases or conditions including coronary heart disease and
 CC other cardiovascular diseases, and sitosterolemia-associated condition
 CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
 CC expression cassette is useful in the production of transgenic non-human
 CC animals. SSG genes and their homologues are useful as tools for a number

CC of applications including diagnosing sitosterolemia and other
 CC cardiovascular disorders, for forensic and paternity determinations,
 CC and for treating any of a large number of SSG associated diseases. The
 CC present sequence is mouse SSG protein variant obtained by replacing
 CC Ile17 with Leu.
 CC Note: The present sequence is not shown in the specification but is
 CC derived from mouse SSG protein referred as SEQ ID NO: 1 (AAEL3308) and
 CC shown in figure 7 of the specification.

XX Sequence 652 AA:

Query Match 82.5%; Score 2742.5; DB 23; Length 652;
 Best Local Similarity 80.2%; Pred. No. 4.7e-280;
 Matches 523; Conservative 64; Mismatches 64; Indels 1; Gaps 1;

OY 1 MGDLISLTPGSGSLQVNRGSSQSLGATATAPER-HSIGIILHASYSHRRPMTDINS 59
 DB 1 MGELPFLSPGARGPHLNRGSLSLGQSVGTTEARSHVLSYSTVSRGPMWNINS 60
 OY 60 CROQWTRQILKDVSLYVESQIIMCIIIGSSGKRTLLDMSGRGAGTFLGEVYVNGRA 119
 DB 61 COQKMDROLKDVSLYIESGQIMCIIIGSSGSKRTLLDMSGRGAGTFLGEVYVNGCE 120
 OY 120 LRREPQDCFSYVLOSLLSLVYRETLHYTALLAIRGNPGSFQKKEAVYAAELISLH 179
 DB 121 LRDDQFQDCFSYVLOSDFSLVYRETLHYTALLAIRGNSADFYKKEAVYMTETLSLH 180
 OY 180 VADRILGNSLGISGERRRYSIAQLQDDKVMFLDEPPTGICOMTNOIVYLLEYLA 239
 DB 181 VADMIGSNFGGISGERRRYSIAQLQDDKVMFLDEPPTGICOMTNOIVYLLEYLA 240
 OY 240 RRRNRIVLTIHOPRSELPQDFPKIAIISFGEILFCPTAPAMLDPFNDGCPPEHSNPD 299
 DB 241 RRDRIYIYTIHQPRSLFQHPFKIAIITYGELFCPTPEMGLFNNCGYPCPREHNPDP 300
 OY 300 FYMDLTSVDQSKEREIEFSKRVQIMESAYKSAICHKTUKNIEMKHILKTLPPWPKTK 359
 DB 301 FYMDLTSVDQSKEREIEFSKRVQIMESAYKSAICHKTUKNIEMKHILKTLPPWPKTK 360
 OY 360 DSPGYSKGLVLRRTYRLVNRKLAIVTRLLONIMGLFLFPTLYRVSNVILKQVQDR 419
 DB 361 DPPGYSKGLVLRRTYRLVNRKLAIVTRLLONIMGLFLFPTLYRVSNVILKQVQDR 420
 OY 420 VGLLQVFGATPTTGLNANVNFVPLRAVSDSDGLYOKQMMLAYALHVPESVAT 479
 DB 421 VGLLQVFGATPTTGLNANVNFVPLRAVSDSDGLYOKQMMLAYALHVPESVAT 480
 OY 480 MIFSSVCYWTGLLHPEVAFGFSALLAPHLIGELFVLGIYONPNIVNSVALLSI 539
 DB 481 VIFSSVCYWTGLLHPEVAFGFSALLAPHLIGELFVLGIYONPNIVNSVALLSI 540
 OY 540 AGVLVSGFLRNIOEMPIFKIISYFTFOKCYSEILVNEFGYGLNFTCGSSNVSTYTNM 599
 DB 541 SGLLIGSGFIRNIOEMPIFKIISYFTFOKCYSEILVNEFGYGLNFTCGSSNVSTYTNM 600
 OY 600 CAPTGOIPIETKTCGATSRFTMNLILYSPALVILGIYVKIKIDHLISR 651
 DB 601 CAITGVQFIETKTCGATSRFTMNLILYSPALVILGIYVKIKIDHLISR 652

RESULT 10
 AA096985
 ID AA096985 standard; Protein: 652 AA.
 AC AA096985;
 DE 30-JUL-2002 (first entry)
 XX Mouse ABCG5 protein.
 DE Mouse ABCG5 protein.
 KM Mouse; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KM atherosclerosis; heart disease; hypersterolemia; Alzheimer's disease.
 XX

OS	Mus sp	Location/Qualifiers
XX	Key	Misc-difference 638..652
XX	FT	/note="Encoded by CTag"
XX	XX	MO200227016-A2.
XX	PD	04-APR-2002.
XX	PR	25-SEP-2001; 2001MO-US29859.
XX	XX	25-SEP-2000; 2000US-235268P.
XX	PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.
XX	PA	(PATE/) PATEL S B.
XX	PI	(DEAN/) DEAN M.
XX	PI	Patel SB, Dean M;
XX	DR	WPI: 2002.416483/44.
XX	DR	N-PSDB: ABR51684.
XX	PT	Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
XX	PT	nucleic acid encoding the polypeptide, useful for treating
XX	PT	sitosterolemia, arteriosclerosis and heart diseases
XX	XX	Example 3; Page 42; 66pp; English.
XX	XX	The present invention relates to a new mammalian ATP-binding cassette
XX	XX	gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
XX	XX	preposition for developing sitosterolemia, arteriosclerosis or heart
XX	XX	disease. The molecules of the invention are also useful for identifying
XX	XX	a compound which alters ABCG5 activity level comprising contacting a cell
XX	XX	culture or mammal which have ABCG5 polypeptide with a compound and
XX	XX	measuring ABCG5 biological activity in the cell culture or in mammal,
XX	XX	where an increase or decrease in ABCG5 biological activity compared to
XX	XX	ABCG5 biological activity in a control cell culture or mammal not
XX	XX	contacted with the compound, identifies a compound that increases or
XX	XX	decreases ABCG5 activity respectively. The cell culture or mammal
XX	XX	comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
XX	XX	ABCG5 biological activity' or level of ABCG5 mRNA, or level of the
XX	XX	polypeptide in a cell culture or mammal is also compared with that of a
XX	XX	second cell culture or mammal comprising a wild type ABCG5 polypeptide.
XX	XX	Stimulation of ABCG5 activity is useful for treating or preventing
XX	XX	hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
XX	XX	disease. The method of the invention is useful for increasing cholesterol
XX	XX	excretion and/or decreasing cholesterol adsorption. The present amino
XX	XX	acid sequence represents the mouse ABCG5 protein of the invention.
XX	XX	Sequence 652 AA;
XX	XX	Query Match 82.3%; Score 2738.5; DB 23; Length 652;
XX	XX	Best Local Similarity 80.1%; Pred. No. 1.2e-279;
XX	XX	Matches 522; Conservative 64; Mismatches 65; Indels 1; Gaps 1
XX	XX	1 MGDLSLTGPGSMGLQVNRGSSSILEGAPATAPED-HISGILTHASYSVHRYPMMWDITS 59
XX	XX	1 MGEPLPFLSPGARGHINRGSLSLSDGSSVCTEARHSLGVLHVSYSNRKGPMMNITS 60
XX	XX	60 CROGTRQILKDVLSYVESGOIMCILLSSGSGKTTLLDMSGRGLRAGCTFLGEVYVNGRA 119
XX	XX	61 COQKMDROIILKDVSYIESGOIMCILLSSGSGKTTLLDAISGLRLARTGTLLBEVEVNGCE 120
XX	XX	120 LRROFOFCFSVYLOSDFLLSLTATRETLHTALTAIRKGNNGSGSQKVEAYEAMTSLSH 179
XX	XX	121 LRROFQDCFSVYLOSDFLLSLTATRETLHTALTAIRKGNNGSGSQKVEAYEAMTSLSH 180
XX	XX	180 VADRLLIGNYSLGISTGERRRYSIAQQLADPPKVMFLPEPTGLGCMATANOIVVLVLETA 239
XX	XX	181 VADQMIGSYNRGSGISGERRRYSIAQQLADPPKVMFLPEPTGLGCMATANOIVVLVLETA 240
XX	XX	240 RRNRIVLTHIQPNSFLQFLDKIAILISFGELIFCGPAEMLDPENDGYPGPCPENSINPD 299

Dd		241	RRORIYIVTTHQPRSELFQHFKIALVITGELVFPGTPEMGLFFNNCGYCPHSPNPFD	3000
Oy		300	FYNDLISVTQSKERELEFSKRVOMTESAYRKSACIKHTLANIERMKHLTKLPVVPETK	3598
Dd		301	FYMDLISVDPSOSHEREIEFYKKRVOMDECAFKESEDIIYKHILEINIERARFLTKLPVVPETK	3600
Oy		360	DSGVSCKLGVLIRRTNRILVNKNKLAVITRLLONIMLFLFPLFRARSVNLGAQIDR	4198
Dd		361	DPEMGKRLGVALLRRVTRNLRMRKQOVIMRLVQNIMGLFIYLLFRQNTLTGAVODR	4200
Oy		420	VGLLYQVFGATPYTGMLNANVNFPLYRAVSDQESODGXOKKOMMLAVALVLPESVAT	4798
Dd		421	VGLLYQLVGATPYTGMLNANVNFPLYRAVSDQESDGLXHKQMHLAVLVLPESVIAT	4800
Oy		480	MIFSSVCYWTGLSHPEVARPFGYSALAPHLIGEFLLTVLLGIYQNPNIYNSVALTLST	5398
Dd		481	VIFSSVCYWTGLGYPEVARPFGYSALAPHLIGEFLLTVLLGIYQNPNIYNSVALTLST	5400
Oy		540	AGVANSGLFNRLNQEMPIPKIITSFTPTKYCSSELYVNERYGALFTGGSSNVSTTNPM	5998
Dd		541	SGLLISSGFIRNLQEMPIPKIITLGYTFPOKCCCELLVNERYGALFTGGGSNTSLNHMP	6000
Oy		600	CAPTQGIQFIKTCPCGATSRETMNFLTIFYSFIPALVILGIYVEKRHDLSR	6518
Dd		601	CALTQGVQFIKTCPCGATSRETMNFLTILGFIPLALVILGIYVFRDYLSR	6520
RESULT 11				
	AAU96986			
ID	AAU96986	standard; protein; 652 AA.		
Ac	AAU96986;			
Dt				
Dt	30-JUL-2002	(first entry)		
Xx				
De	Rat ABCG5 protein.			
Kw	Rat; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;			
Kw	arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease.			
Os	Rat sp.			
Xx				
Pd				
Pn	W0200227016-A2.			
Fn				
En	04-APR-2002.			
Xx				
Pf	25-SEP-2001: 2001MO-US29859.			
Px				
Pr	25-SEP-2000: 2000US-235268P.			
Xx				
Pa	(USSH) US DEPT HEALTH & HUMAN SERVICES.			
Pa	(PATE/) PATEL S B.			
Pa	(DEAN/) DEAN M.			
Pi	Patel SB, Dean M;			
Xx				
Dr	WPI: 2002-416483/44.			
Dr	N-PDSB; ABR51686.			
Xx				
Pt	Novel mammalian ATP-binding cassette gene 5 polypeptide, and the			
Pt	nucleic acid encoding the polypeptide, useful for treating			
Pt	sitosterolemia, arteriosclerosis and heart diseases -			
Xx				
Ps	Example 3; Page 45; 66pp. English.			
Xx				
Cc	The present invention relates to a new mammalian ATP-binding cassette			
Cc	gene 5 (ABCG5) polypeptide. The invention is useful for identifying a			
Cc	predisposition for developing sitosterolemia, arteriosclerosis or heart			
Cc	disease. The molecules of the invention are also useful for identifying			
Cc	a compound which alters ABCG5 activity level comprising contacting a cell			
Cc	culture or mammal which have ABCG5 polypeptide with a compound and			
Cc	measuring ABCG5 biological activity in the cell culture or in mammal,			

CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity, respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the rat ABCG5 protein of the invention.

Sequence 652 AA:

Query Match 82.0%; Score 2727.5; DB 23; Length 652;
 Best Local Similarity 79.4%; Pred. No. 1.8e-278;
 Matches 518; Conservative 68; Mismatches 65; Indels 1; Gaps 1;

QY 1 MGDLSLTPGSGMGLQVNRGSSSLGAPATAPAP- HSLGILHASYSVSHRVPMDITSC 59
 DB 1 MGEPLPSPEGARGHNNGSSOSLSEGSVTGSEARHSIGVNLVSFVSNRVGMWNITKS 60
 QY 60 RCOQWTRQILKDVSLYVESGOIMCTLGSSGSKTLLDAMSGRGRAGTFLGEVYVNGRA 119
 DB 61 COQKDKILKDVSLYISGOTMCTLGSSGSKTLLDAMSGRGRAGTFLGEVYVNGRA 120
 QY 120 LRREGQDCFSYVLSQDPLSLSLYVETLHYTALAIRNGNGSFQKVEAMALSLSH 179
 DB 121 LRBDQFQDCVSLQSDVFLSLYVETLHYTALAIRNGNGSFQKVEAMALSLSH 180
 QY 180 VADRLIGNSLGGISGERRRYSIAQLLODPKVMLEDEPTTGDCMTANOIVLLVEIA 239
 DB 181 VADQMGVNGSGISGERRRYSIAQLLODPKVMLEDEPTTGDCMTANOIVLLVEIA 240
 QY 240 RNRNIVYTIHQPRESELFPOLFKAILSGELIPGCTPAEMDFNDCGCPDEHSNPD 299
 DB 241 RNRNIVYTIHQPRESELFPOLFKAILSGELIPGCTPAEMDFNDCGCPDEHSNPD 300
 QY 300 FYMDLTVSDTOSREHEIETSKRVOMIESAYKSAICHTLKNIERMKHLKTLPMVPEKTK 359
 DB 301 FYMDLTVSDTOSREHEIETSKRVOMIESAYKSAICHTLKNIERMKHLKTLPMVPEKTK 360
 QY 360 DSPGFESLGLVLRVTNLRNKLAVTRLLQNLIMGLFLFVLRVRSNVKGAIDR 419
 DB 361 NPGFECRLGLVLRVTNLRNKLAVTRLLQNLIMGLFLFVLRVRSNVKGAIDR 420
 QY 420 VGLLYQFYGATPYTGMLNAVNLFPYLRAVSDOESODGLYOKOMMLAVALHVPESVAT 479
 DB 421 VGLLYQFYGATPYTGMLNAVNLFPYLRAVSDOESODGLYOKOMMLAVALHVPESVAT 480
 QY 480 MIFSSVCYWTGLHPEVARFGYSFSAALLAPHLIGFLLVLLGIYONPNINYSVALLSI 539
 DB 481 VIFSSVCYWTGLHPEVARFGYSFSAALLAPHLIGFLLVLLGIYONPNINYSVALLSI 540
 QY 540 AGVYNGSEFLNIQEMPIPKIISFTFQKCSSELVNVENYGLNFTGSSNVSYTNP 599
 DB 541 SGLLGSGEFLNIQEMPIPKIISFTFQKCSSELVNVENYGLNFTGSSNVSYTNP 600
 QY 600 CAFTQGIQFIKTCGATSRFTMNFLLIXSFLPALVILGIYVFKIRHLLISR 651
 DB 601 CSMTQGIQFIKTCGATSRFTMNFLLIXSFLPALVILGIYVFKIRHLLISR 652

RESULT 12

AA096991 standard; ProteIn; 408 AA.

AA096991;

30-JUL-2002 (first entry)

DE Human ABCG5 mutant R408X protein sequence.
 XX Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW mutant; mutelin.
 XX Homo sapiens.
 OS Synthetic.
 FT Key Location/Qualifiers
 FT Misc-difference 408
 FT /note= "Wild-type protein truncated at this position"
 XX WO200227016-A2.
 XX 04-APR-2002.
 XX 25-SEP-2001; 2001WO-US29859.
 XX 25-SEP-2000; 2000US-235268P.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX (PATE/) PATEL S B.
 XX (DEAN/) DEAN M.
 XX Patel SB, Dean M;
 XX WPI: 2002-416483/44.
 XX Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 PS Claim 10; Page -; 66pp; English.

The present invention relates to a new mammalian ATP-binding cassette
 gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 predisposition for developing sitosterolemia, arteriosclerosis or heart
 disease. The molecules of the invention are also useful for identifying
 a compound which alters ABCG5 activity level comprising contacting a cell
 culture or mammal which have ABCG5 polypeptide with a compound and
 measuring ABCG5 biological activity in the cell culture or in mammal,
 where an increase or decrease in ABCG5 biological activity compared to
 ABCG5 biological activity in a control cell culture or mammal not
 contacted with the compound, identifies a compound that increases or
 decreases ABCG5 activity, respectively. The cell culture or mammal
 comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 mutant R408X protein of the
 CC invention.
 CC Note: This sequence is not shown in the specification but is derived
 CC from the wild-type human ABCG5 protein (AA096984) given on pages 35-36
 CC of the specification.

Sequence 408 AA:

Query Match 62.6%; Score 2081; DB 23; Length 408;
 Best Local Similarity 100.0%; Pred. No. 1.7e-210;
 Matches 408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGDLSLTPGSGMGLQVNRGSSSLGAPATAPAP- HSLGILHASYSVSHRVPMDITSC 60
 DB 1 MGEPLPSPEGARGHNNGSSOSLSEGSVTGSEARHSIGVNLVSFVSNRVGMWNITSC 60
 QY 61 RCOQWTRQILKDVSLYVESGOIMCTLGSSGSKTLLDAMSGRGRAGTFLGEVYVNGRA 120
 DB 61 RCOQWTRQILKDVSLYVESGOIMCTLGSSGSKTLLDAMSGRGRAGTFLGEVYVNGRA 120

QY 121 RREFODCFSTVLAQSDTLSSLVRETLHYTALLAIRKNGSGFQKVEAVMAELSLSHV 180
 CC |||||||
 Db 121 RREFODCFSTVLAQSDTLSSLVRETLHYTALLAIRKNGSGFQKVEAVMAELSLSHV 180
 QY 181 ADRLIGNSLGGISTGERRRVSIAAQLDPRKMLFDEPTTGDCMTANQIVLVVELAR 240
 CC |||||||
 Db 181 ADRLIGNSLGGISTGERRRVSIAAQLDPRKMLFDEPTTGDCMTANQIVLVVELAR 240
 QY 241 RNRIVLTIHOPRSELFPOLDKIALISFGELFCGTPAEMLDFFNDCGYPCEHSNPPDF 300
 CC |||||||
 Db 241 RNRIVLTIHOPRSELFPOLDKIALISFGELFCGTPAEMLDFFNDCGYPCEHSNPPDF 300
 QY 301 YMDLTSVDQSKEREIETSKRVOMIESAYKSAICHKTLKNIERKHLKTLPMVPFKTD 360
 CC |||||||
 Db 301 YMDLTSVDQSKEREIETSKRVOMIESAYKSAICHKTLKNIERKHLKTLPMVPFKTD 360
 QY 361 SPGVFSKLGVLRRVTNVLNKNLAIVITRLQNLINGLFLFVLAVR 408
 CC |||||||
 Db 361 SPGVFSKLGVLRRVTNVLNKNLAIVITRLQNLINGLFLFVLAVR 408
 Db 361 SPGVFSKLGVLRRVTNVLNKNLAIVITRLQNLINGLFLFVLAVR 408
 RESULT 13
 AAU96987 ID AAU96987 standard; Protein; 340 AA.
 AC AAU96987;
 XX
 DT 30-JUL-2002 (first entry)
 DE Hamster ABCG5 protein.
 XX
 KM Hamster; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KM arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease.
 OS Citellinae sp.
 XX
 PN NC0200227016-A2.
 XX
 PD 04-APR-2002.
 XX
 PF 25-SEP-2001; 2001MO-US29859.
 XX
 PR 25-SEP-2000; 2000US-235268P.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATEL) PATEL S B.
 PA (DEAN/) DEAN M.
 XX
 PI Patel SB, Dean M;
 DR WPI: 2002-416483/44.
 DR N-PSDB; ABK51687.
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 XX
 PS Example 3; Page 46; 66pp; English.
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the

CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the hamster ABCG5 protein of the invention.
 XX
 SQ Sequence 340 AA:
 Query Match 41.7%; Score 1387.5; DB 23; Length 340;
 Best Local Similarity 78.6%; Pred. No. 2,2-137;
 Matches 264; Conservative 39; Mismatches 32; Indels 1; Gaps 1;
 QY 98 AMSGRAGFTLGEVYVNGRALRREFODCFSTVLAQSDTLSSLVRETLHYTALLAIR 157
 CC |||||||
 Db 1 AISGRRLRTGTLEGEVFNNGRELRRDQFODCFSTVLAQSDVFLSLVRETLHYTALLAIR 60
 QY 158 RGNFGSFQKVEAVMAELSLSHVADRILGNTSLGGISTGERRRVSIAAQLDPRKMLFD 217
 CC |||||||
 Db 61 SSSDFYDKKVEAVMAELSLSHVADRILGNTSLGGISTGERRRVSIAAQLDPRKIMFD 120
 QY 218 EPTTGDCMTANQIVLVVELARNRRLVLTIIHOPRSELFPOLDKIALISFGELFCGTP 277
 CC |||||||
 Db 121 EPTTGDCMTANQIVLVVELARNRRLVLTIIHOPRSELFPOLDKIALISFGELFCGTP 180
 QY 278 AEMLDFFNDCGYPCEHSNPPDFYMDLTSVDQSKEREIETSKRVOMIESAYKSAICHK 337
 CC |||||||
 Db 181 EEMLDFFNDCGYPCEHSNPPDFYMDLTSVDQSKEREIETSKRVOMIESAYKSAICHK 240
 QY 338 TKNIERKHLKTLPMVPFKTOSPGVFSKLGVLRRVTNVLNKNLAIVITRLQNLING 397
 CC |||||||
 Db 241 TLENIERTKHLKTLPMVPFKTOSPGVFSKLGVLRRVTNVLNKNLAIVITRLQNLING 300
 QY 398 LFLFVLAVRNSVNLKGAIDRVGLTKQFGATPYT 433
 CC |||||||
 Db 301 LFLFVLAVRQNDILKGAIDRVGLTSM-SAPFRT 335
 RESULT 14
 AAU96988 ID AAU96988 standard; Protein; 243 AA.
 AC AAU96988;
 XX
 DT 30-JUL-2002 (first entry)
 DE Human ABCG5 mutant R243X protein sequence.
 XX
 KM Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KM arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KM mutant; mutain.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 FT Key Location/Qualifiers
 FT Misc-difference 243 /note- "Wild-type protein truncated at this position"
 XX
 PN NC0200227016-A2.
 XX
 PD 04-APR-2002.
 XX
 PF 25-SEP-2001; 2001MO-US29859.
 XX
 PR 25-SEP-2000; 2000US-235268P.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATEL) PATEL S B.
 PA (DEAN/) DEAN M.
 XX
 PI Patel SB, Dean M;

XX DR WPI: 2002-416483/44.
 XX PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 XX PS Claim 13: Page -: 66pp: English.
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 mutant R243X protein of the
 CC invention.
 CC Note: This sequence is not shown in the specification but is derived
 CC from the wild-type human ABCG5 protein (AA096984) given on pages 35-36
 CC of the specification.
 XX SQ Sequence 243 AA:
 Query Match 37.1%; Score 1234; DB 23; Length 243;
 Best Local Similarity 100.0%; Pred. No. 2.1e-121;
 Matches 243; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 MCDLSLPFGSGMGLQVNRGSSQSLGAPATAPEPHSILILASYSVSHRPMPMDITSC 60
 DB 1 MDSLSTLPFGSGMGLQVNRGSSQSLGAPATAPEPHSILILASYSVSHRPMPMDITSC 60
 OY 61 RQOMRROLIKDYSIVVESQIMCIGSSSGKTTLLDAMSGRLAGTFLGVEYVNGRAL 120
 DB 61 RQOMRROLIKDYSIVVESQIMCIGSSSGKTTLLDAMSGRLAGTFLGVEYVNGRAL 120
 OY 121 RREQFODCFYSYVLOSSTLLSLVRETHYTLALAIRGNPSSFOKRYEAVNAEISLSHY 180
 DB 121 RREQFODCFYSYVLOSSTLLSLVRETHYTLALAIRGNPSSFOKRYEAVNAEISLSHY 180
 OY 181 ADRLIGNYSLGISTGERRRYSIAAQLQDPKVMLEDEPTTGLDCTANQIVYLVELAR 240
 DB 181 ADRLIGNYSLGISTGERRRYSIAAQLQDPKVMLEDEPTTGLDCTANQIVYLVELAR 240
 OY 241 RNR 243
 DB 241 RNR 243
 RESULT 15
 ID AAB41856 standard; Protein: 144 AA.
 XX AAB41856;
 AC AAB41856;
 XX 08-FEB-2001 (first entry)
 DT 08-FEB-2001 (first entry)
 XX Human ORFX ORF1620 polypeptide sequence SEQ ID NO:3240.
 DE Human ORFX ORF1620 polypeptide sequence SEQ ID NO:3240.
 XX Human: open reading frame: ORFX; detection: cytosstatic; hepatotropic;
 KW vulnereary; antiposoriatic; antiparkinsonian; nootropic; neuroprotective;

KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX OS Homo sapiens.
 XX PN WO200058473-A2.
 XX PD 05-OCT-2000.
 XX PF 31-MAR-2000; 2000MO-US08621.
 XX PR 31-MAR-1999; 99US-0127607.
 XX PR 02-APR-1999; 99US-0127636.
 XX PR 05-APR-1999; 99US-0127728.
 XX PR 30-MAR-2000; 2000US-0540763.
 XX PA (CURA-) CURAGEN CORP.
 XX PI Shinkets RA, Leach M.
 XX DR WPI: 2000-602362/57.
 XX DR N-PSDB: AAC76065.
 XX PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX PS Claim 11: Page 2444; 5507pp: English.
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytosstatic; hepatotropic; vulnereary;
 CC antiposoriatic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX SQ Sequence 144 AA:
 Query Match 21.7%; Score 722; DB 21; Length 144;
 Best Local Similarity 99.3%; Pred. No. 1.1e-67;
 Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 357 KTKDSPGVFSKLGVLRRVTNLRVNRKLAIVTTRLLQNLIMGLFLFPLRVRSVNLGAI 416
 DB 1 KTKDSPGVFSKLGVLRRVTNLRVNRKLAIVTTRLLQNLIMGLFLFPLRVRSVNLGAI 60
 OY 417 ODRVGLLYQFVGATPYTGMLNAVNLFPYLRAVSQDSQDGLYQKWMQMLAVALVLPFSV 476
 DB 417 ODRVGLLYQFVGATPYTGMLNAVNLFPYLRAVSQDSQDGLYQKWMQMLAVALVLPFSV 476

Wed Jun 11 09:51:14 2003

us-09-989-981a-6.rag

Page 13

Db 61 QDRVGLXQFYGAPPTGMLNAVNLFPVLRASDOSODGLYOKWOMLAVALHVLPSV 120

Oy 477 VATMIFSSVCYTGLGHPVARFG 500

Db 121 VATMIFSSVCYTGLGHPVARLG 144

Search completed: June 11, 2003, 09:12:15
Job time : 43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: June 11, 2003, 09:04:42 ; Search time 15 Seconds

(Without alignments)
1800.073 Million cell updates/sec

Database: US-09-989-981a-6

Sequence: 1 MGDLSLTPGSGMGLQVNRG.....PALVILGIVFKIRDLISR 651

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: SwissProt.40.4

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	3326	100.0	651	ABG5_HUMAN
2	2738.5	65.2	1	ABG5_MOUSE
3	2727.5	65.2	1	ABG5_RAT
4	698	21.0	673	ABG8_MOUSE
5	697	21.0	673	ABG8_HUMAN
6	691	20.8	672	ABG2_HUMAN
7	676.5	20.3	655	ABG2_MOUSE
8	627	18.9	1294	TOH5_YEAST
9	623	18.7	677	WHIT_YEAST
10	621	18.7	1049	ADP1_YEAST
11	607.5	18.3	695	WHIT_MOUSE
12	602.5	18.1	687	WHIT_DROME
13	596.5	17.9	678	ABG1_HUMAN
14	591	17.8	679	WHIT_CERCA
15	578.5	17.4	646	ABG4_HUMAN
16	573	17.2	709	WHIT_ANOL
17	569.5	17.1	666	ABG1_MOUSE
18	561.5	16.9	598	YPC3_CAEEL
19	547	16.4	666	SCRT_DROME
20	517.5	15.6	610	YOSC_CAEEL
21	454.5	13.7	675	BROW_DROME
22	435	13.1	668	BROW_DROVI
23	427	12.8	1499	CDR3_CANAL
24	406.5	12.2	1333	YNG2_YEAST
25	403.5	12.1	1564	PDR4_YEAST
26	399.5	12.0	1530	BFR1_SCHPO
27	398	11.8	1501	CDR3_CANAL
28	392.5	11.8	1529	PDRF_YEAST
29	391	11.5	650	ABG3_MOUSE
30	383.5	11.5	1501	CDR1_CANAL
31	382	11.5	1511	PDR5_YEAST
32	382	11.5	1511	PDR5_YEAST
33	374	11.2	1511	PDR5_YEAST

ALIGNMENTS

RESULT 1	ABG5_HUMAN	STANDARD:	PRT:	651 AA.
AC	Q9H222			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DE	ATP-binding cassette, sub-family G, member 5 (Sterol-1).			
GN	ABG5.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Carnivora; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND VARIANT GLD-604.			
RC	TISSUE=Liver.			
RA	Medline-20553648; PubMed-11099417;			
RA	Berge R.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,			
RA	Willeverich J., Shan B., Barnes R., Hobbs H.H.,			
RA	Accumulation of dietary cholesterol in the liver of mice with			
RA	mutations in adjacent ABC-transporters.			
RL	Science 290:1771-1775(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A., VARIANTS SITOSTEROLEMA H-389; H-419 AND P-419,			
RP	AND VARIANT E-604.			
RC	TISSUE=Liver.			
RA	Medline-20578753; PubMed-11138003;			
RA	Lee M.-H., Lu K., Hazard S., Yu H., Shulien S., Hidaka H., Kojima H.,			
RA	Dean M., Patel S.B.,			
RA	Identification of a gene, ABCG5, important in the regulation of			
RA	dietary cholesterol absorption.			
RL	Nat. Genet. 27:79-83(2001).			
RN	[3]			
RP	REVIEW.			
RA	Medline-21474438; PubMed-11590207;			
RA	Schmitz G., Langmann T., Helmerl S.,			
RA	Role of ABCG1 and other ABCG family members in lipid metabolism.			
RL	J. Lipid Res. 42:1513-1520(2001).			
RN	[4]			
RP	VARIANTS SITOSTEROLEMA Q-146; H-389; P-419; H-419 AND S-550, AND			
RP	VARIANT E-604.			
RA	Medline-21344600; PubMed-11453359;			
RA	Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,			
RA	Stalenhoef A.F.H., Meeklenburg T., Bjorkhem I., Bruckert E.,			
RA	Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,			
RA	Patel S.B.,			
RA	Two genes that map to the STRL locus cause sitosterolemia: genomic			
RA	structure and spectrum of mutations involving sterol-1 and			
RA	sterolin-2, encoded by ABCG5 and ABCG8, respectively.			
RL	Am. J. Hum. Genet. 69:278-290(2001).			
CC	-1- FUNCTION: Transporter that appears to play an indispensable role			
CC	in the selective transport of the dietary cholesterol in and out			
CC	of the enterocytes and in the selective sterol excretion by the			
CC	liver into bile.			
CC	-1- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to			

CC -1- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to

Db	1	MDLISLSPFGSMGLQVNNGSOSLSLEGARATAPERPSSGLIHLNASTSVSHRYRRWMDITSC	60
QY	61	ROOMTRQIILKDVSLVYESGOIMCIISSGSGGKTTLLDAMSGRLGRAETGLEYVNGRAL	120
Db	61	ROOMTRQIILKDVSLVYESGOIMCIISSGSGGKTTLLDAMSGRLGRAETGLEYVNGRAL	120
QY	121	REBQVODCSYVLQSTLSSLTVBKTLHTATALLAIRGNPGSGFQKKEAVMAELSHV	180
Db	121	REBQVODCSYVLQSTLSSLTVBKTLHTATALLAIRGNPGSGFQKKEAVMAELSHV	180
QY	181	ADRLIGNSLGISTGERRRVSIQAOLDDPYMLDEPTTGIDCKTANOIVYLVEIAR	240
Db	181	ADRLIGNSLGISTGERRRVSIQAOLDDPYMLDEPTTGIDCKTANOIVYLVEIAR	240
QY	241	RNRIVYLTIHOPRSELEFOLFRIATALISFEGELIFCGTPAEMLDFENCCGCPCEPESHNPDE	300
Db	241	RNRIVYLTIHOPRSELEFOLFRIATALISFEGELIFCGTPAEMLDFENCCGCPCEPESHNPDE	300
QY	301	YMDLTSDVTOSKERELETSKRVOMIESAVKKAICHKTLKNIERMKHLTLPVWPCKTD	360
Db	301	YMDLTSDVTOSKERELETSKRVOMIESAVKKAICHKTLKNIERMKHLTLPVWPCKTD	360
QY	361	SPGVSKSLGVLLRRYTRNRYRKLAVITPRLQNLINGLELLFFVLRVRSNZLKGAIQDRY	420
Db	361	SPGVSKSLGVLLRRYTRNRYRKLAVITPRLQNLINGLELLFFVLRVRSNZLKGAIQDRY	420
QY	421	GLIYFVGATPYTGMILNANVLEFPVLRASVDSQSDGLYOKOMMLATYALHVPFSVATM	480
Db	421	GLIYFVGATPYTGMILNANVLEFPVLRASVDSQSDGLYOKOMMLATYALHVPFSVATM	480
QY	481	IFSSVCYMTGLAHEVARGVFSALLAPHLIGEFTLVILGIYONPNIVNSVVALLSIA	540
Db	481	IFSSVCYMTGLAHEVARGVFSALLAPHLIGEFTLVILGIYONPNIVNSVVALLSIA	540
QY	541	GVLVSGSLRNQENPIPKIISYTFPOKYCEIIVYNEFYGLNFCGSSNVSYTTPNC	600
Db	541	GVLVSGSLRNQENPIPKIISYTFPOKYCEIIVYNEFYGLNFCGSSNVSYTTPNC	600
QY	601	AFTQGIQIIEKTCPGATSRFTMNFILIXSIFALYALIGIVYKIRDHLSR	651
Db	601	AFTQGIQIIEKTCPGATSRFTMNFILIXSIFALYALIGIVYKIRDHLSR	651

```

RT mutations in adjacent ABC transporters.
RL Science 290:1771-1775(2000).
CC -1- FUNCTION: Transporter that appears to play an indispensable role
CC in the selective transport of the dietary cholesterol in and out
CC of the enterocytes and in the selective sterol excretion by the
CC liver into bile.
CC -1- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
CC ABCG8 along a pathway regulating dietary-sterol absorption and
CC excretion (by similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- TISSUE SPECIFICITY: Expressed in the intestine and, at lower
CC level, in the liver.
CC -1- INDUCTION: Upregulated by cholesterol feeding. Possibly mediated
CC by the liver X receptor/retinoic X receptor (LXR/RXR) pathway.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AF312713; AAC53097.1;
DR MGD; MG1:1351659; Abcg5.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
DR ATP-binding; Glycoprotein; Transmembrane; Transport.
FT DOMAIN 1 385 1 (POTENTIAL).
FT TRANSMEM 386 406 1 (POTENTIAL).
FT TRANSMEM 407 422 2 (POTENTIAL).
FT TRANSMEM 423 443 2 (POTENTIAL).
FT TRANSMEM 444 463 3 (POTENTIAL).
FT TRANSMEM 464 484 3 (POTENTIAL).
FT TRANSMEM 485 504 4 (POTENTIAL).
FT TRANSMEM 505 525 4 (POTENTIAL).
FT TRANSMEM 526 529 5 (POTENTIAL).
FT TRANSMEM 530 550 5 (POTENTIAL).
FT TRANSMEM 551 622 6 (POTENTIAL).
FT TRANSMEM 623 643 6 (POTENTIAL).
FT DOMAIN 644 652 6 (POTENTIAL).
FT NP_BIND 87 94 6 (POTENTIAL).
FT CARBOHYD 410 410 6 (POTENTIAL).
FT CARBOHYD 585 585 6 (POTENTIAL).
FT CARBOHYD 592 592 6 (POTENTIAL).
SQ SEQUENCE 652 AA; 73244 MW; 80CE37ADCC19771E CRC64;

Query Match 82.3%; Score 2738.5; DB 1; Length 652;
Best Local Similarity 80.1%; Pred. No. 1.3e-183;
Matches 522; Conservative 64; Mismatches 65; Indels 1; Gaps 1;

OY 1 MGDLSLTPGSMGLOVNRGSSSLGGAATAPAP-HSGILHASTSVSHRRPMDITS 59
DB 1 MELPELSPGAPGPIINGSLSLGQSVGTGTEAHSGVHYVSYVSNRYGPMWNIRS 60
OY 60 CROQMTQILKDVSLTVESGOINCLIGSSGSGKTTLLDMSCGLGAGFLFVYVNGRA 119
DB 61 CQKMDROLKDVSLTVESGOINCLIGSSGSGKTTLLDMSCGLGAGFLFVYVNGRA 120
OY 120 LRRDPQDFSTVLQSDTLSSLTVEHLYTALLAIRNGNGSFQKTEAVAAELSLSH 179
DB 121 LRRDPQDFSTVLQSDTLSSLTVEHLYTALLAIRNGNGSFQKTEAVAAELSLSH 180
OY 180 VADRLGNSTLGGISGERRRVSIAQLQDPKVMLEPPTGGLDQMTANQIVLLVELA 239
DB 181 VADRLGNSTLGGISGERRRVSIAQLQDPKVMLEPPTGGLDQMTANQIVLLVELA 240

OY 240 RRRRIYVTHIOPRSELEFQDKIAISFGLFCGTPALMDPENCDCPCPEHSNPD 299
DB 241 RRRRIYVTHIOPRSELEFQDKIAISFGLFCGTPALMDPENCDCPCPEHSNPD 300
OY 300 FYMDLTSVDTQSKERELETSKRVQMESAYKSAICHTLKNTERMKHLTLPMPVEPKT 359
DB 301 FYMDLTSVDTQSKERELETSKRVQMECAKESDIYHKLLENERRAYLTLPVPEPKT 360
OY 360 DSGVSKGLVLLRRTYRNRYKKAIVITRLLONLNGLEFLAFVYLRNSVULKAIQDR 419
DB 361 DPGMFGKGLVLLRRTYRNRYKKAIVITRLLONLNGLEFLAFVYLRNSVULKAIQDR 420
OY 420 VGLIYFVGATPYTGMNAVNLPPVLRVAVSDOSOGLYOKOMMLAYALHYVPSVAT 479
DB 421 VGLIYFVGATPYTGMNAVNLPPVLRVAVSDOSOGLYOKOMMLAYALHYVPSVAT 480
OY 480 MIFSSVCYWTGLHPEVARGFSAALLAHLGEEFLVLLGIYONPNIVNSVVALSTI 539
DB 481 VIFSSVCYWTGLHPEVARGFSAALLAHLGEEFLVLLGIYONPNIVNSVVALSTI 540
OY 540 AGVYSGFLRNQOEMPPIPKIISTYTPQKCYSELVYNEFYGLNFTCGSSNVSTYTPM 599
DB 541 SGLIISGFLRNQOEMPPIPKIISTYTPQKCYSELVYNEFYGLNFTCGSSNVSTYTPM 600
OY 600 CAPTQGIQFIETKCPGATSRFTNMFILVSEIPALVILGIVPKIRDHLSR 651
DB 601 CATIQVOFTEKCPGATSRFTNMFILVSEIPALVILGIVPKIRDHLSR 652

RESULT 3
ABG5_RAT STANDARD; PRT; 652 AA.
ID ABG5_RAT
AC 099PE7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
GN ABG5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Small Intestine;
RX MEDLINE=20578753; PubMed=11138003;
RA Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,
RA Bean M., Patel S.B.,
RT "Identification of a gene, ABCG5, important in the regulation of
RT dietary cholesterol absorption."
RL Nat. Genet. 27:79-83(2001).
CC -1- FUNCTION: Transporter that appears to play an indispensable role
CC in the selective transport of the dietary cholesterol in and out
CC of the enterocytes and in the selective sterol excretion by the
CC liver into bile.
CC -1- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
CC ABCG8 along a pathway regulating dietary-sterol absorption and
CC excretion (by similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AF312714; AAC53098.2;
DR InterPro: IPR003593; AAA_ATPase.

```

InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transporter; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KM ATP-binding; Glycoprotein; Transmembrane; Transport.
 FT DOMAIN 1 385
 FT TRANSMEM 386 406
 FT DOMAIN 407 422
 FT TRANSMEM 423 443
 FT DOMAIN 444 463
 FT TRANSMEM 464 484
 FT DOMAIN 485 504
 FT TRANSMEM 505 525
 FT DOMAIN 526 529
 FT TRANSMEM 530 550
 FT DOMAIN 551 624
 FT TRANSMEM 625 645
 FT DOMAIN 646 652
 FT NP_BIND 87 94
 FT CAROXYD 585 585
 FT CAROXYD 592 592
 SQ SEQUENCE 652 AA; 73342 MW; 4DA2FE2BAB0DD59 CRC64;
 Query Match 82.0%; Score 2727.5; DB 1; Length 652;
 Best Local Similarity 79.4%; Pred. No. 7.7e-183;
 Matches 518; Conservative 68; Mismatches 65; Indels 1; Gaps 1;
 1 MDDLSTLPPGSGMGLOVNGSSSLEGPATAPAP-HSLGILHASYSHRRPMDITS 59
 1 MCELFLPSPGARGHNRNNGSSSLEGGVTSERHSAGLVNVSANRGPWMNKS 60
 60 CROQMTROLKNDVSLYVESGQIMCTIGSSGKTTLLDAMSGRLNAGTFLEGVNNGRA 119
 61 COOKMDRLKNDVSLYIESGQIMCTIGSSGKTTLLDASRLNAGTFLEGVNNGRA 120
 120 LRRBPPOCFVYVSDTLSSLYRRETLATALLAIRNGPSQOKVEAVALSLSH 179
 121 LRRDPPOCFVYVSDTLSSLYRRETLATALLAIRNGPSQOKVEAVALSLSH 180
 180 VADRLIGNTSGISGTERRRYSIAQLDQPKVLFDEPTTGLOMANTAVLVLELA 239
 181 VADQMGNTSGISGTERRRYSIAQLDQPKVLFDEPTTGLOMANTAVLVLELA 240
 240 RNRRIIVLTIRPSELEQLEFKIALISFGLIFCGTPAEMLDFFNDGCGPCPEHSNPD 299
 241 RNRRIIVLTIRPSELEQLEFKIALISFGLIFCGTPAEMLDFFNDGCGPCPEHSNPD 300
 300 FYMDLTVDYQCKEKEIEISKRVQMEISYKKSALCHTKLKNIEKMKLTKLPMPFKTK 359
 301 FYMDLTVDYQCKEKEIEISKRVQMEISYKKSALCHTKLKNIEKMKLTKLPMPFKTK 360
 360 DSPGVFSKLGVLRRVTRNLVNRKLAIVTRLLONTIMGFLFVLIRVRSNVLKGAIDR 419
 361 NRPGRFCKLGVLRRVTRNLVNRKLAIVTRLLONTIMGFLFVLIRVRSNVLKGAIDR 420
 420 VGLLQFGADPYTGMLNANVLPVLRVAVSDQSDGGLYOKQMMALAVLHLPFSVAT 479
 421 VGLLQFGADPYTGMLNANVLPVLRVAVSDQSDGGLYOKQMMALAVLHLPFSVAT 480
 480 MEESVCYWTGLHEVARFGFSALLAPHLIGELTLYLVLLGYNPNRYNSVALLSI 539
 481 MEESVCYWTGLHEVARFGFSALLAPHLIGELTLYLVLLGYNPNRYNSVALLSI 540
 540 AGVLVSGFLNRNIQEMPIPKIISYTFQKCYSELLVNEFGNFTGGSSVASTPDM 599
 541 SGLLGGFIRNIEEMPIPKIISYTFQKCYSELLVNEFGNFTGGSSVASTPDM 600
 600 CAFTGIGPIEKTGPGATSRFTMNLILYSFIPALVILIGIVFKIRDLISR 651
 601 CSMTGIGPIEKTGPGATSRFTMNLILYSFIPALVILIGIVFKIRDLISR 652

RESULT 4
 ID ABG8_MOUSE
 AC 09D8M0;
 DT 15-JUN-2002 (rel. 41, Created)
 DT 15-JUN-2002 (rel. 41, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 8 (SterolIn-2).
 GN ABG8.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 [1]
 SEQUENCE FROM N.A. (ISOPFORMS 1 AND 2).
 RC STRAIN=C57BL/6; TISSUE=Liver;
 RX MEDLINE=21344600; PubMed=11452359;
 RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
 RA Ose L., Stalenhoef A.F.H., Miettinen T., Bjorkhem I., Bruckert E.,
 RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Sirtavasta A.K.,
 RA Patel S.B.;
 RT "Two genes that map to the STSL locus cause sitosterolemia: genomic
 RT structure and spectrum of mutations involving sterolIn-1 and
 RT sterolIn-2, encoded by ABG5 and ABG8, respectively.";
 RL Am. J. Hum. Genet. 69:278-290(2001).
 [2]
 SEQUENCE FROM N.A. (ISOPFORM 1).
 RC STRAIN=C57BL/6; TISSUE=Liver;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shilagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann M., Gaasterland T., Gissi C., King B., Kochiya H.,
 RA Kuehl M., Lewis S., Matsuo Y., Mikaido I., Pesole G., Quackenbush J.,
 RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Blake J., Bokil D., Fujino M., Aono H., Baldarini R., Barsh G.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincin S., Hill D., Hofmann M., Hume D.A., Kanuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whitaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 [3]
 TISSUE SPECIFICITY, AND INDUCTION.
 RP MEDLINE=20553648; PubMed=11099417;
 RX Beyer K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
 RA Kwieterovich P., Shan B., Barnes R., Hobbs H.H.;
 RT "Accumulation of dietary cholesterol in sitosterolemia caused by
 RT mutations in adjacent ABC transporters.";
 RL Science 290:1771-1775(2000).
 CC -!- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.
 CC -!- SUBUNIT: May form heterodimers with ABG5 or be tightly coupled to
 CC ABG5 along a pathway regulating dietary-sterol absorption and
 CC excretion (by similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
 CC produced by alternative splicing.
 CC -!- TISSUE SPECIFICITY: Expressed in the intestine and, at lower
 CC level, in the liver.
 CC -!- INDUCTION: Upregulated by cholesterol feeding, possibly mediated
 CC by the liver x receptor/retinoid x receptor (LXR/RXR) pathway.
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABG8 (WHITE)
 CC SUBFAMILY.

```

CC      -1- CAUTION: Seems to have a defective ATP-binding region.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: AF324495; AAR64079.1;
DR      EMBL: AK004871; BAB23630.1;
DR      MGI: 1914720; Abcg8.
DR      InterPro: IPR003439; ABC_transporter.
DR      Pfam: PF00005; ABC_tran.1.
DR      ProDom: PD000006; ABC_transporter.1.
DR      PROSITE: PS00211; ABC_TRANSPORTER.1.
KW      Glycoprotein; Transmembrane; transport; Alternative splicing.
FT      DOMAIN 1 413
FT      TRANSSEM 414 434 1 (POTENTIAL).
FT      DOMAIN 435 447 1 (POTENTIAL).
FT      TRANSSEM 448 468 2 (POTENTIAL).
FT      DOMAIN 469 496 1 (POTENTIAL).
FT      TRANSSEM 497 517 3 (POTENTIAL).
FT      DOMAIN 518 526 4 (POTENTIAL).
FT      TRANSSEM 527 547 5 (POTENTIAL).
FT      DOMAIN 548 569 5 (POTENTIAL).
FT      TRANSSEM 570 590 6 (POTENTIAL).
FT      DOMAIN 591 639 6 (POTENTIAL).
FT      TRANSSEM 640 660 6 (POTENTIAL).
FT      DOMAIN 661 673 6 (POTENTIAL).
FT      CARBOHYD 677 677 673 N-LINKED (GLCNAC...) (POTENTIAL).
FT      VARSPLIC 379 377 MISSING (IN ISOFORM 2).
SQ      SEQUENCE 673 AA; 75995 MW; 78012611A5D2559 CRC64;

Query Match 21.08; Score 698; DB 1; Length 673;
Best Local Similarity 28.7%; Pred. No. 3,2e-41;
Matches 194; Conservative 133; Mismatches 264; Indels 84; Gaps 18;

OY      11 GSGGLGVNRSOSSLECAPATAPEPHSGLLHASYSHYR-----PKWD-I 57
DB      14 GYLVDASQGLDLSL-----FSSSDNSLVFTYSGSNTLEPRLTYQVDIASQVPMPEQL 69
OY      58 TSCROOQMTROI-----LKDVSLVYSGSIOIMCIISSGSGTTLIDAMSGRLGRAGTF 109
DB      70 AQKRIPIRSHSSSDSCSELGRINISFVRSQGMALITSSCGRASILDVITGR-GHGKM 128
OY      110 -LGEVTVNGRRLRREQPDFQDFSVYLSQDLSLVVETLHYTALLAI-RRGNPGSFQK 167
DB      129 KSGQIWINQPTPTQVLRKCAVHAHRQHDQLPLNLTAFIAQMLPRTFSGAQRDKR 188
OY      168 VEAVMAELSHVADLLIGNYSLGISTGERBRVSTAQQLQDPKWLAEDEPTTGIDCMT 227
DB      189 VEDVIAELRLROCCANRVTGYRVSQGGERRRVSIGVOLAMPGLIILDEPTSGIDSTF 248
OY      228 ANQIVLLVLELARRNRIIVLTIHQPSLELFQDLKALISFEGILFCGTPAEMLDFENDC 267
DB      249 AHNLTVTTLRLKAGNKLVLISLHQPDSDFRLDLVLMISGPITVLGAQAQWVOYFTST 308
OY      288 GIPCEPHSNPDFYMDLQSVQSKERELETSKRVOMIESAVYKSA-----ICHTLKN 341
DB      309 GHCPKRYSNPADFYVDLTSDIRSKEREVATVEKASIALFLEKVGDFDPMKKEAKE 368
OY      342 IERMKRLKTLPMWFEFTKDS-----FGVESKLVILARVTRNLVYRNKLAVITRLLQN 393
DB      369 LMTSTHTVSLTL-----TQDPLDCGTAVELFGMIEQFSTLIRKQISNFRDIPFTLLIGSEA 424
OY      394 LIMGLELFFVLVRSNVLKGAIQDVRGLLYGVGATPYTGMANVNLFPVLVAVDQES 453
DB      425 CLMSLIGFLYGHGAKOL--SPMDPALLFAGIALIPRVLLIDVYSKCHSESMLYTEL 482
OY      454 QDGLYQKQOMLAVLALVLPFSVVAITMIFSSVCYWTILGLHPEVARGYFSALLAHLIG 513

```

```

DB      483 EDGLYTAGPYFAKILGELPERCAVVIYAMPYMLNLRPVELF-----LL--HLL 534
OY      514 EEL-----TLVLGIVQNPVI-VNSVVALISLGVGSGFLNRIOEMPIPEKISTFT 566
DB      535 VMLVVFQCCFRFMAASAMLPFHMSSEFCNALVNSFYLTAGFMINDMLIVPAMISKLS 594
OY      567 FOKYCEILVYNEFYGLNFT--CGSSNVSTTNMCAFTQGIQTEKTCPATSRFTNMF 624
DB      595 FLRMCFSGLMQIYNGHLYTQIGNFTFPTSLGDM-----ISAMDLS 637
OY      625 LILVSPALVITGI 639
DB      638 HPLVAIY--LIVIGI 650

RESULT 5
ABCG8_HUMAN STANDARD; PRT; 673 AA.
ID      ABCG8_HUMAN
AC      Q9H221;
DT      15-JUN-2002 (Rel. 41, Created)
DT      15-JUN-2002 (Rel. 41, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      ATP-binding cassette, sub-family G, member 8 (Sterol-in-2).
GN      ABCG8.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX      NCBI_Taxid=9606;
RN      [1]
RP      SEQUENCE FROM N.A. VARIANTS SITOSTEROLEMITA T-231; Q-263; R-574 AND
RP      R-596, AND VARIANT C-54.
RX      MEDLINE-20533646; PubMed-11099417;
RA      Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
RT      Kvitserovich P., Shan B., Barnes R., Hobbs H.H.;
RT      "Accumulation of dietary cholesterol in sitosterolemia caused by
RL      mutations in adjacent ABC transporters.";
RL      Science 290:1771-1775(2000).
RN      [2]
RP      SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), VARIANTS SITOSTEROLEMITA H-184;
RP      T-231; Q-263; H-405; P-501; S-543; P-572; E-574; R-574; R-596 AND
RP      F-570 DEL, AND VARIANTS H-19; C-54; K-238; V-259; K-400; R-575 AND
RP      A-632.
RC      TISSUE-Liver;
RX      MEDLINE-21344600; PubMed-11452359;
RA      Lee K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
RA      Ou L., Stalenhoef A.F.H., Miettinen T., Bjorkhem I., Bruckert E.,
RA      Pandey A., Brewer H.B. Jr., Salen G., Dean M., Silvestava A.K.,
RA      Patel S.B.;
RT      "Two genes that map to the SRSF locus cause sitosterolemia: genomic
RT      structure and spectrum of mutations involving sterolin-1 and
RT      sterolin-2, encoded by ABCG5 and ABCG8, respectively.";
RL      Am. J. Hum. Genet. 69:278-290(2001).
RN      [3]
RP      REVIEW.
RX      MEDLINE-21474438; PubMed-11590207;
RA      Schmitz G., Langmann T., Helmerl S.;
RT      "Role of ABCG1 and other ABCG family members in lipid metabolism.";
RL      J. Lipid Res. 42:1513-1520(2001).
CC      -1- FUNCTION: Transporter that appears to play an indispensable role
CC      in the selective transport of the dietary cholesterol in and out
CC      of the enterocytes and in the selective sterol excretion by the
CC      liver into bile.
CC      -1- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to
CC      ABCG5 along a pathway regulating dietary-sterol absorption and
CC      excretion.
CC      -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC      Produced by alternative splicing. Isoform 1 (shown here) and 2; are
CC      detected in approximately 10% of the cDNA clones.
CC      -1- TISSUE SPECIFICITY: Strongly expressed in the liver, lower levels
CC      in the small intestine and colon. Detectable in a wide variety of
CC      human tissues.
CC      -1- DISEASE: Defects in ABCG8 are a cause of sitosterolemia, also

```


FT	VARIANT	19	19
FT	/FTID=VAR_012250.		
FT	VARIANT	54	54
FT	/FTID=VAR_012251.		
FT	R -> H (IN STROSTROLEMIA)		
FT	P -> T (IN STROSTROLEMIA)		
FT	/FTID=VAR_012252.		
FT	E -> K.		
FT	/FTID=VAR_012253.		
FT	A -> V.		
FT	/FTID=VAR_012254.		
FT	A -> Q (IN STROSTROLEMIA)		
FT	/FTID=VAR_012255.		
FT	R -> O (IN STROSTROLEMIA)		
FT	/FTID=VAR_012256.		
FT	VARIANT	263	263

[illegible]

RESULT 6
ABG8_RAT STANDARD: PRT: 672 AA.

AC P58428: 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 8 (Sterol-2).
GN ABCG8.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN: Sprague-Dawley.
RX MEDLINE-21344600; PubMed-11452359;
RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
RA Ose L., Stalenhoef A.F.H., Miettinen T., Bjorkhem I., Bruckert E.,
RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Silvestre A.K.,
RA Patel S.B.;
RT "Two genes that map to the SRSF locus cause sitosterolemia: genomic
RT structure and spectrum of mutations involving sterol-1 and
RT sterol-2, encoded by ABCG5 and ABCG8, respectively.";
RL Am. J. Hum. Genet. 69:278-290(2001).
CC -1- FUNCTION: Transporter that appears to play an indispensable role
CC in the selective transport of the dietary cholesterol in and out
CC of the enterocytes and in the selective sterol excretion by the
CC liver into bile.
CC -1- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to
CC ABCG5 along a pathway regulating dietary-sterol absorption and
CC excretion (By similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
CC produced by alternative splicing.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
CC SUBFAMILY.
CC -1- CAUTION: Seems to have a defective ATP-binding region.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE351785; AAK8431.1;
CC InterPro: IPR003593; AAA_ATPase.
CC InterPro: IPR003439; ABC_transporter.
CC Pfam: PF00005; ABC_tran. 1.
CC SMART: SM00382; AAA. 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW Glycoprotein; Transmembrane; Transport; Alternative splicing.
FT DOMAIN 1 412
FT TRANSMEM 413 433
FT TRANSMEM 434 446
FT DOMAIN 447 467
FT TRANSMEM 468 495
FT TRANSMEM 496 516
FT TRANSMEM 517 525
FT TRANSMEM 526 546
FT TRANSMEM 547 568
FT TRANSMEM 569 589
FT TRANSMEM 590 628
FT TRANSMEM 629 649
FT DOMAIN 650 672
FT CARBOHYD 618 618
FT VARSPICD 376 376
FT SEQUENCE 672 AA: 75878 MW: 60889735FA4C36F CRC64;
Query Match 20.8%; Score 691; DB 1; Length 672;
Best local Similarity 28.3%; Pred. No. 9.8e-41;

Matches 189; Conservative 126; Mismatches 264; Indels 88; Gaps 15;

OY 23 SLECAPATADEPHSIGILHASYSVSHRVR-----PW-----WDITSC 60
DB 21 SSLDQSVSSSDNSLFTYTGSGSNLEVRDLTYQVDMAQVPMPEQALQKLPWRSGS 80
OY 61 RQMTROLKQVSLYESGQIMCIGSSGSKTLLDMSGRGRLGACGAEVYVNGRAL 120
DB 81 QDSMDLGI-RNLSFKVRSGOMALIGSACGRALDITGRDGGKSKSGQIMNGPS 139
OY 121 RREOFQDFSVYLDSTSLTYRETHYATLAIKRGNGSF-----QKVENYMAEL 175
DB 140 TPQLQKCVAHVRQDDLLPMLTYRETHYATLAIKRGNGSF-----QKVENYMAEL 195
OY 176 SLSHVADRLNIGSYLGGISTGERRRVSIQAQDQPKVMEDEPTTGIDCTANQIVYL 235
DB 196 RLROCAVTRVGNVTVRGVSGGERRRVSIGVQLLNPGLILDEPTSGIDSTAINVRTL 255
OY 236 VELARRNRIVVLTTHORSELPOLFDKAILISFELLIPCTPAEMLDPFNDGYPCEHS 295
DB 256 SRLAKGRNLYLISLHOPRSDIFRLEPDLVLLMTSGPIYLGAQHMVQFTSIGYCPERS 315
OY 296 NPPEFYMDLTSVDQSKRETEETSKRVOMIESAKKSA-----ICKTKLNIEMRHLK 349
DB 316 NPADFYDLSIDRSKQEVATMEKARLLALFLERKVGQDFDLKRAEAKSLD----TG 371
OY 350 TLPWVPEKTKDS-----PGVFSKGLVLRVTRNLRNKLAVITRLQNLINGFL 401
DB 372 TYAVSQTLQDFTNGCTAELRPMIQOFTTLRLQISDFRDLPLFLTHGAEACMSLIG 431
OY 402 FETVLRVSNVILKAIQDRVCLLYPGATPTTGALNAVNEFPVLRVSDQSGLYQKW 461
DB 432 FLYYGADKPL--SEPMMAALLFMIGALIPFNVLIDVYSKHSRSLYELLEDGLTAG 489
OY 462 QMMLAYALVLPFVSVAVTMIFSSVCYTWLGHPEVARGFYSALLNPHLIGEL----- 516
DB 490 PYFAKVLGELPEKCAVYIIGMPYIWTNLRP-----GPELFLHPFLMLLVFCC 541
OY 517 -TLVLGIVQNPNI-VNSVVALISAGLVGSLRNIOEMPIFKIISYTFQKYSCEI 574
DB 542 RYMLAASAMLPFRHMSFCCNALNYSFYLAGMINLMIVPAMISKMSFLRMKFSG 601
OY 575 LYVNEFGLNFTGSSSVSTTNWCAFTGIGTETGCG--ATSFTNFIILYSFIP 632
DB 602 LMQIQFNHGYTTOIGMLTFESV-----PGDAVMTAMDLSNHLVAYI- 643
OY 633 ALVILGI 639
DB 644 -LVIYGI 649

RESULT 7
ABG2_HUMAN STANDARD: PRT: 655 AA.

AC Q9UNQ0: Q95374; Q9UNQ0; Q9B773;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 2 (Placenta-specific ATP-
DE binding cassette transporter) (Breast cancer resistance protein).
GN ABCG2 OR ABCP OR BCRP OR BCRP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE: Placenta;
RX MEDLINE-99065313; PubMed-9850061;
RA Allikmets R., Schriml L.M., Hutchinson A., Romano-Spica V., Dean M.;
RT "A human placenta-specific ATP-binding cassette gene (ABCP) on
RT chromosome 4q22 that is involved in multidrug resistance.";
Cancer Res. 58:5337-5339(1998).

[2]
RN SEQUENCE FROM N.A.
RC TISSUE-Breast cancer:
RA MEDLINE-99080071; PubMed-9861027;
RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,
RA Ross D.D.;
RA "A multidrug resistance transporter from human MCF-7 breast cancer
RA cells";
RA Proc. Natl. Acad. Sci. U.S.A. 95:15665-15670(1998).
RN [3]
RA ERRATUM.
RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,
RA Ross D.D.;
RA Proc. Natl. Acad. Sci. U.S.A. 96:2569-2569(1999).
RN [4]
RA SEQUENCE FROM N.A.
RA Rake K., Tsukahara S., Sugiyama T., Asada S., Ishikawa E., Tsuruo T.,
RA Sugimoto Y.;
RA "Breast cancer resistance protein constitutes a 140-kDa complex as a
RA homodimer";
RA Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
RA SEQUENCE OF 198-655 FROM N.A.
RA TISSUE-Placenta;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiraori A., Sudo H.,
RA Wagaetsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Watanabe S., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Yamamoto J., Makamatsu A., Nakamura Y., Nagahara K., Masuo Y.,
RA Niimura K., Iwayanagi T.;
RA "NEO human cDNA sequencing project";
RA Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RA REVIEW.
RA MEDLINE-21474438; PubMed-11590207;
RA Schmitz G., Langmann T., Helmetl S.;
RA "Role of ABCG1 and other ABCG family members in lipid metabolism";
RA J. Lipid Res. 42:1513-1520(2001).
RN [7]
RA FUNCTION: XENOBIOTIC TRANSPORTER THAT APPEARS TO PLAY A MAJOR ROLE
RA IN THE MULTIDRUG RESISTANCE PHENOTYPE OF A SPECIFIC MCF-7 BREAST
RA CANCER CELL LINE. WHEN OVEREXPRESSED, THE TRANSFECTED CELLS BECOME
RA RESISTANT TO MITOXANTHRONE, DAUNORUBICIN AND DOXORUBICIN, DISPLAY
RA DIMINISHED INTRACELLULAR ACCUMULATION OF DAUNORUBICIN, AND
RA MANIFEST AN ATP-DEPENDENT INCREASE IN THE EFFLUX OF RHODAMINE 123.
RA SUBCELLULAR LOCATION: Integral membrane protein (Probable).
RA -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
RA SUPERFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
DR EMBL: AF103796; AAD09188.1;
DR EMBL: AF098951; AAC97367.1;
DR EMBL: AB056867; BAB39212.1;
DR EMBL: AK002040; BAA92050.1;
DR EMBL: HGNC:74; ABCG2.
DR MIM: 603756;
DR InterPro: IPR0003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF00005; ABC_tran.1.
DR ProDom: PD000006; ABC_transportr.1.
DR SMART: SM00382; AAA.1.
DR PROSITE: PS00211; ABC_TRANSPORTER; FALSE_NEG.
DR ATP-binding; Transmembrane; Transport.
DR DOMAIN 1 395 CYTOPLASMIC (POTENTIAL).
DR TRANSMEM 396 416 POTENTIAL.
DR DOMAIN 417 428 EXTRACELLULAR (POTENTIAL).

FT	TRANSMEM	429	449	POTENTIAL.
FT	DOMAIN	450	477	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	478	498	POTENTIAL.
FT	DOMAIN	499	506	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	507	527	POTENTIAL.
FT	DOMAIN	528	535	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	536	556	POTENTIAL.
FT	DOMAIN	557	630	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	631	651	POTENTIAL.
FT	DOMAIN	652	655	CYTOPLASMIC (POTENTIAL).
FT	NP_BIND	80	87	ATP (POTENTIAL).
FT	CARBOHYD	418	418	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	557	557	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	596	596	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	24	24	V -> A (IN REF. 2 AND 4).
FT	CONFLICT	166	166	E -> Q (IN REF. 1).
FT	CONFLICT	208	208	F -> S (IN REF. 5).
FT	CONFLICT	315	316	MISSING (IN REF. 2).
FT	CONFLICT	482	482	R -> T (IN REF. 2).
FT	SEQUENCE	655 AA;	72343 MW;	89A6D3511DC5CE0 CRC64;

Query Match 20.3%; Score 676.5; DB 1; Length 655;
Best Local Similarity 29.0%; Pred. No. 9,8e-40;
Matches 181; Conservative 137; Mismatches 251; Indels 55; Gaps 18;

21 SOSLEGAPATAP---EPHSGLHNSYSVSRVRRMMDITSCROQWTRQILKDYSLYE 77
13 SGNRTGFRATYSNDLKATBEAVLSFNHCIRVKLSSFLPCRKEVELLSNNGIMK 72
78 SQOIMCITLSSGSGRTLLDAMSGRLAGRTFGEYVNGRRLRREOFQDCFSYVLSQDT 137
73 PG-LNAILDPTGGKSSLLDVLAAKRDPSG-LSGVLING-APRANPCNSGYVQDDV 129
138 LLSLTVRETLHTYLLAIRGNPG-SFOKVEAYVAELSLSHVADRILGNTSLGISTG 196
130 VMGTLVRENLOFSALRLATMTNHEKERNIRYEEIGLDKVDKSGTQFINGVSG 189
197 EERRVIAQLLODPRVLMFDEPTGLOCMNTQNOVIVLYELARRNRYVLTIRPRSL 256
190 EKRKSIGMELTDDSLFLDEPTGLDSSANAVALLLRKSKGGRITTFIHOPRSY 249
257 FQLEPKIALSPGELIFGTPAEMLDFFNDGCPCEPHSNPPEFMDLTSVPTQ----SK 313
250 FKLFDLSLTLASGRMLFHPGPAQEGALGYESAGYHEANNNPADFLDIINGSTVALNR 309
313 EEE-----IETSKR-----VQMBEAYKSAICHT-----LKNIERKHLKTLPMVP 356
310 EEDFKATELIEPSKODKPLIEKIAEIVYNSPYKETKAKELHQLSGEKKKKTIVKEISY 369
357 KTKDSPGVSKLGVLLRVTNRLVANKLAVITRLONLIMGLF--LEFVLRVSNVLKG 414
370 TT-----SCHOLRWYSKSKFKNLNPQASIAQIIVYVGLVIGAITFGKLNDSF---- 421
415 AIDRVGILLYOFVATPYTGMLNANVLEPVYLAASDOSDGLYOKRMOMLALVAL-HYLP 473
422 GIONRAGLYF-FLTNQCFSSVAVELVEYKEKLFHRYISGYRVSSYPIGLKSLDLP 480
474 FSVVATMIESSVCYTTLGHEPVARFGFSALLAPHLIGELTVLYLGIYONPRTYNSV 533
481 MRMLDSILFTCIYVFMGLKPRADAFVMMETLM---WVAYSASSMALAIAGGSVSA 537
534 VALLSIAGV--LVGSGFLRNIOEMPIPKIISYFFQKCGSEILVNVNEGYLNGCGSSN 591
538 TLIMKICVFMIMISGLLVNLTITASMLSMLOYSIPRYGTALOHNEFLDQNCPCG--- 594
592 VSVTTPNCACTOGIQTETKCPG 615
595 LNAIGNNPCNYA-----TCTG 610

RESULT 8
YOH5_YEAST
ID YOH5_YEAST STANDARD; PRT; 1294 AA.

Query Match	18.9%;	Score 627;	DB 1;	Length 1294;
Best Local Similarity	31.7%;	Pred. No. 6.7e-36;		
Matches 181;	Conservative 106;	Mismatches 228;	Indels 56;	Gaps 19

QY	179	HVAARLIGNSLIGSITGSTERRRVSLAAOLDPKVMEFDEPTTGIDCMTANOIVLVAYL	238
Db	825	HCENNIIGNEVAKGISGGKRRVTVGVLLDPLLLDDEPTSGISDSTSAITLLEKRL	884
QY	239	AR-NRRIVLTTHIOPRSELFOLEFKIALIS-FGELIFCPTPEAMIDPFNDGCGPCPEHSN	296
Db	885	CREGKFTIITIHQPRSLFRFGVLLLASGRFAFGSDDEMIATYTELGYNCPSFTN	944
QY	297	PFEDYMDLTSVDTSKEREIETSKRQOMIESAYKKAICHTLKNIEKKHLKTLPMVPE	356
Db	945	VADFDLLDSVNTONEONEISSRAREVEKILSAMK-----ANMDN-ESISPPISEK	994
QY	357	KTKDSPGFSKLGVLIRVRFTNLY-----RUKIAVITRLLQNLINGLLEL	401
Db	995	QOYQOESFEFTYSEFVRK-PANVLAIYAVVNRKROFTTTRRSDDSIIMARIAPGIGVIFA	1055
QY	402	FEVLAVSNVLKGAIODRVGLLYXFGVATPYTGMLNAVNLEPVLAVADQESODGLYORX	461
Db	1054	LEFAPVKNHNT--SISNRLLGAQST-ALYFVGMIGNLACQPTENDYFEEENOVYGLA	1110
QY	462	QMLATYALHVPESVATMTMISVCYTWTGLHPYARARGYSALLAPHLI--GEFLTL	518
Db	1111	PEFLATYMTLEPLISALASVLAVFTVLACG-PRTA-GNPFATVYCSFIYTCCEERLGI	1167
QY	519	VLLGIYONPN-IVNSVALLISIAVLVSGFLRNIOEMPIFKIISYTPQKYSGEILVY	577
Db	1168	MTNPFEPHPCGVVACISITISIGOMSLML-----GMSRLKGNYNLNPVGYTSMIITN	1222
QY	578	NEFYG-LNFTC--GSSNVSVTTNPMCAFTQG	605
Db	1224	FAFPGNLKLCEDGKNSDGT-----CEFRANG	1250

RESULT 9

WHIT_LUCCU	STANDARD:	PRT;	677 AA.
ID	WHIT_LUCCU		
AC	005360;		
DT	01-FEB-1995 (Rel. 31, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	White protein.		
GN	W.		
OS	Lucilia cuprina (Greenbottle fly) (Australian sheep blowfly).		
OC	Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;		
OC	Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;		
OC	Muscomorpha; Oestroidea; Calliphoridae; Lucilia.		
OX	NCBI_Taxid=7375;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE-97087158; Pubmed-8933176;		
RA	Garcia R.L., Perkins H.D., Howells A.J.;		
RT	"The structure, sequence and developmental pattern of expression of		
RT	the white gene in the blowfly <i>Lucilia cuprina</i> ."		
RL	Insect Mol. Biol. 5:251-260(1996).		
RN	[2]		
RP	SEQUENCE OF 490-584 FROM N.A.		
RP	MEDLINE-90264941; Pubmed-1971656;		
RA	Elizur A., Vacek A.T., Howells A.J.;		
RT	"Cloning and characterization of the white and topaz eye color genes		
RT	from the sheep blowfly <i>Lucilia cuprina</i> ."		
RL	J. Mol. Evol. 30:347-358(1990).		
CC	-FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM		

CC NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
CC CELLS RESPONSIBLE FOR EYE COLOR.
CC
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial


```

FT CAROHD 971 971 N-LINKED (GLCNC...) (POTENTIAL).
SQ SEQUENCE 1049 AA: 117231 MW: ABC9CE54BCDF6A3 CRC64;
Query Match 18.7%; Score 621; DB 1; Length 1049;
Best Local Similarity 28.6%; Pred. No. 1,3e-35;
Matches 196; Conservative 111; Mismatches 223; Indels 156; Gaps 22;

OY 68 ILKDVSLVYESGOIMCILLSSGSGGTTLLDAMSGRAGTFLGEVYVGRALRREOFD 127
DB 405 VLNEISGIYKPCOILALMGSGAGKTTLLDILAM-RTKHGHSGSIRKVGISMOKDSFSK 463
OY 128 CESYLOSDTLLSSTVETLHTYLLAIRGNPGSPQK---VEAVALSLSHVADR 184
DB 464 IIGFDODDFLLPLTVEETVNSALRLPKAL--SFEAKKARVYKVELEIRIIDIKDR 521
OY 185 IGNYSLSGISTGERRRRSIAOQLDPKYMFEDEPTTGIDCMTANOIVLVELAR-RNR 243
DB 522 IGNEDRIGISGERKRSVLAELVPLFDEPTSGIDANANNVTECTVRLSDYNR 581
OY 244 IVTLTHQPSSELPFDKIALISFGLIFCGTPAEMLDFNDCGYPCEPHSNPEFYMD 303
DB 582 TLVLSTHQPNSNIFELFKVLVLSKGEVYSGNAKVSSEFLNEGICPDNYNIADYLD 641
OY 304 LT-SVDQSKEREI----- 316
DB 642 ITFPAQPKRRRRIRNISDLEAGTDTNDINTIHOTFTSDGTQREMAHLAHRDEIR 701
OY 317 -----ETSKRYOMIESAYKKSALCHIKLKNIERM----- 345
DB 702 SLNDEDEVEGTGDRGATEIDNTKLLHDKKOSVYAEISOETEELSEGEDESNVLAN 761
OY 346 KHLATLPNVPKTKDSQPSVSKGLVLRVTRNLRNKLAVITRLLQMLMGFLFLFVL 405
DB 762 GDLEPT-----GOOSAGFLOQLSLNRSRFRNMNRNKLKLLGNLTLILSLFGTLY 814
OY 406 RVRNSVLKGAIQDRVGLY---QRYGATPYTGMLNAVLFPVLAHSVDSQSDGLYQWQ 462
DB 815 NV-SNDISG-FQNRMGLEFFLTYGFTFTGL-----SSFALERIIFIKERSNNYSP-- 866
OY 463 MHLAYAL-----HYLPSSVATMIESSVCYTICLHPVAPRGFSALLAPHLIGEFLT 517
DB 867 --LAVYSKIMSEVPLRVPLVPLISLIVPTGSLNMDNF-FKICIGILLFLFNGLISLE 923
OY 518 LVLLGIY---ONPNIIVSVALLSIAGVLSGFLRNQOEMP-IPKIIISYFTOKYCE 573
DB 924 ILTIGIIFEDLNSIISLVLL---GSLFSGLEFNKNTNNAFKILNFSFYIAYE 980
OY 574 ILVVRNF-----YGLNFTGSSSNVSTTNPMCAFTOGIOFEKTCPGATSRTMNF 624
DB 981 SLLINEKYLMLKRRKGLNI-----EVPATILSTFGF 1014
OY 625 LIUSFIPALVIGI--VVKIRIDL 648
DB 1015 -VQNLVFDIKITALENVFLIMGL 1039

RESULT 11
WHIT ANOGA STANDARD; PRT: 695 AA.
AC Q27256; Q17006;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE White protein.
GN W.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Culicoidae; Anopheles.
OX NCBI_TaxID=7165;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=Suakoko / G3;

RX MEDLINE-96423158; PubMed-8825759;
RA Besansky N.J., Bedell J.A., Benedict M.Q., Mukabayire O., Hlilfiker D.,
RA Collins F.H.;
RT "Cloning and characterization of the white gene from Anopheles
RT gambiae."
RL Insect Mol. Biol. 4:217-231(1995).
CC -1- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
CC NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
CC CELLS RESPONSIBLE FOR EYE COLOR.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U29486; AAC46995.1; -
CC EMBL: U29485; AAC46994.1; -
CC EMBL: U29484; AAC47423.1; -
CC InterPro: IPR003593; AAA_Arase.
CC InterPro: IPR003439; ABC_transporter.
CC InterPro: IPR005284; Pigment_permease.
CC Pfam: PF00005; ABC_tran; 1.
CC Prodom: PM000006; ABC_transporter; 1.
CC SMART; SM00382; AAA; 1.
CC TIGRfams: TIGR00955; 3a01204; 1.
CC PROSITE: PS00211; ABC_TRANSPORTER; 1.
CC KW pigment; ATP-binding; Transmembrane; Transport.
CC NP_BIND 133 140
CC NP_BIND 288 295
CC TRANSMEM 444 464
CC TRANSMEM 474 494
CC TRANSMEM 524 544
CC TRANSMEM 552 572
CC TRANSMEM 581 601
CC TRANSMEM 669 689
CC CAROHD 472 472
CC CAROHD 645 645
CC CONFLICT 100 100
CC CONFLICT 691 693
CC SRS -> IAR (IN REF. 1; AAC47423).
CC SEQUENCE 695 AA: 77218 MW: EEB9517239B2961 CRC64;

Query Match 18.3%; Score 607.5; DB 1; Length 695;
Best Local Similarity 28.4%; Pred. No. 6,9e-35;
Matches 170; Conservative 124; Mismatches 208; Indels 97; Gaps 17;

OY 58 TSCROQWTRQ-----ILKDVSLVYESGOIMCILLSSGSGGTTLLDAMSGRGRAGT 108
DB 96 TLRNCCTRORKDENRKRHLKNYGVAKSGELLAVMSSGKTTLLNLAAR-SPGV 154
OY 109 FLGEVYV---NGRALRROQDCESTYVLSOTLSSLTRETALTAIRGNPGSF- 164
DB 155 KISPAVALNGVPNAQOLRACAYVQDDLFPSLTREHLLEQAMLRMGRDPAVAK 214
OY 165 OKRYEAVVAELSLSHVADRILGNYS-LGISTGERRRRSIAOQLDPKYMFEDEPTTG 223
DB 215 QHRQVEVLESLVKACADTITGAGGRILGSLGGEKKRLAFSETLTDHLLCOEPTGL 274
OY 224 DCMTANOIVLVELARNRNRLVTLTHQPSSELPFDKIALISFGLIFCGTPAEMDF 283
DB 275 DSFMAHSLVGLKGMAMKKTITLTHQPSSELPFDKILVLAGRAFLGSPQSAEF 334
OY 284 FNDGICPCPEHSNPEFDYMDLTSVDQSKEREIETSKRYOMIESAYKKSACHTLKNIE 343
DB 335 FSGIQCPCPNYNADRYVQMLAI--APAEACRCMIRKICOSFAVSPILAREVLEPAS 391
OY 344 RMKHLKTLPMVPEFTKDSQPSVSKL-GV-----LLRRTNLRNKLAVI 387
DB 392 -----VAGKMGDEPIYLDQVEGSGTGRYSMMWTOFYCITLMSWLSVLKDPMLVK 441

```


OY 388 TRILQWLMGLFLFLVLRNVLGAIDRVGL-----LYQVGCATPTTGMNAVL 441
 DB 442 VRLIQTAMAVATLI-----GSIFYGOVLDDGVNMNINSLFLFLNMTFQVAVIN 493
 OY 442 FPLRLAVSDQESDDGLYOKKQMLALALHLVPSVATMFSSVCWTGLHEVARFG 501
 DB 494 FSAELPFLERKSRSLRYRDVYFELGRTIETLPLFLIVPPEFTSITPMICL-----RTG- 547
 OY 502 FSNALLAPHLIGELFLVLGLIYQNP-----IVNSVNLSTA-----GVLCGS 547
 DB 548 -----ATHTL-----TLFLYITLVANVSTSGVILISCASSISMAISVGPVPIPLIFG 598
 OY 548 FLRNIGDMPILPEKIIISYFTFOKXCEILVNEFYGL-----NFTCGSSNVSVTT 596
 DB 599 FLNLSASVPAVYFXYLSLWFRYANBALINOMSTVVDGELIACRANVTCPSREILLET 657

RESULT 12

WHIT_DROME

ID WHIT_DROME STANDARD; PRT; 687 AA.

AC P10090; G9V3A2; G9XY33;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-NOV-1991 (Rel. 20, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE White protein.

GN W OR EG:BACN33B1.1 OR CG2759.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_Taxid=7227;

RN (1)

RP SEQUENCE FROM N.A.

RC TISSUE-Head;

RX MEDLINE=90221897; PubMed=2109311;

RA Pepling M., Mount S.M.;

RT Sequence of a cDNA from the *Drosophila melanogaster* white gene.;

RL Nucleic Acids Res. 18:1633-1633(1990).

RN (2)

RP SEQUENCE FROM N.A.

RX MEDLINE=8513465; PubMed=6084717;

RA O'Hare K., Murphy C., Lewis R., Rubin G.M.;

RT DNA sequence of the white locus of *Drosophila melanogaster*.;

RL J. Mol. Biol. 180:437-455(1984).

RN (3)

RP SEQUENCE FROM N.A.

RX MEDLINE=21100348; PubMed=11156992;

RA Lukatsovich T., Asztalos Z., Awano W., Baba K., Kondo S., Niva S., Yamamoto D.;

RT "Dual-tagging gene trap of novel genes in *Drosophila melanogaster*.";

RL Genetics 157:727-742(2001).

RN (4)

RP SEQUENCE FROM N.A.

RC STRAIN-Berkeley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amaratunga P.G., Scherer S.E., Li P.W., Hoskins R.F.,

RA George R.A., Lewis S.E., Richards M., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Chapple M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Bailew R.M., Basu A., Baxendale J., Bayraktiroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borikova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,

RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Gang N.S., Gelbart W.M., Glasser K.,

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., Moberg S.,
 RA Merklow G., Mlishina N.V., Moberg A.C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of *Drosophila melanogaster*.";
 RA Science 287:2185-2195(2000).

[5] SEQUENCE FROM N.A.

RC STRAIN-Oregon-R;

RX MEDLINE=20196011; PubMed=10731137;

RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,

RA Barrall B.G., Ferraz C., Vidal S., Brun C., Demallies J., Borikova D.,

RA Dreano S., Gloux S., Lelaure V., Motter S., Galibert F., Borikova D.,

RA Pipana B., Katatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,

RA Papagiannakis G., Spanos L., Cox S., Madano E., de Pablos B.,

RA Modolell J., Peter A., Schoettler P., Werner M., Moutriotti F.,

RA Belnert N., Dowe G., Schaefer U., Jaekle H., Bucheton A.,

RA Callister D.M., Campbell L.A., Darlamtsou A., Henderson N.S.,

RA McMillan P.J., Salles C., Tait E.A., Valentini P., Saunders R.D.C.,

RA Glover D.M.;

RT "From sequence to chromosome: the tip of the X chromosome of *D.*RT *melanogaster*.";

RL Science 287:2220-2222(2000).

RN [6]

RP SEQUENCE OF 224-331 FROM N.A.

RX MEDLINE=89339145; PubMed=2503416;

RA Teatle R.G., Belote J.M., McKewon M., Baker B.S., Howells A.J.;

RT Cloning and characterization of the scarlet gene of *Drosophila*RT *melanogaster*.;

RL Genetics 122:595-606(1989).

-1- FUNCTION: PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM NECESSARY

FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT CELLS

RESPONSIBLE FOR EYE COLOR. WHITE DIMERIZE WITH BROWN FOR THE

TRANSPORT OF GUANINE AND WITH SCARLET FOR THE TRANSPORT OF

TRYPTOPHAN.

-1- SUBCELLULAR LOCATION: Integral membrane protein.

-1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.

-1- This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>or send an email to license@sib-sib.ch).

EMBL: X51749; CAA36038.1;

EMBL: X02974; CAA26716.1;

EMBL: AB028139; BAA78210.1;

EMBL: AB004425; AAF45826.1;

EMBL: AL133506; CAB65847.1;

EMBL: X76202; CAA53795.1;

PIR: S07263; FYFW.

PIR: S08635; S08635.

Flybase: FBgn0003996; w.

InterPro: IPR003593; AAA_Arpase.

InterPro: IPR003439; ABC_transportr.

DR InterPro: IPR005284; Pigment_permease.
DR Pfam: PF00005; ABC_tran. 2.
DR ProDom: P0000006; ABC_transporter. 1.
DR SMART: SMO0382; AAA. 1.
DR TIGRfam: TIGR00955; 3a01204.1.
DR PROSITE: PS00211; ABC_TRANSPORTER. 1.
KV Pigment; ATP-binding; Transmembrane; Transport.
FT NP_BIND 130 137 ATP (BY SIMILARITY).
FT TRANSMEM 435 453 POTENTIAL.
FT TRANSMEM 465 485 POTENTIAL.
FT TRANSMEM 515 533 POTENTIAL.
FT TRANSMEM 542 563 POTENTIAL.
FT TRANSMEM 576 594 POTENTIAL.
FT TRANSMEM 659 678 POTENTIAL.
FT CONFLICT 25 29 GDSGA -> LIFEIPYHCRVAD (IN REF. 2 AND 3).
FT CONFLICT 49 49 L-> R (IN REF. 4 AND 5).
FT CONFLICT 335 371 VAOCPNINPDETFVQVLAIVPGRIEISDRRIARIC -> ITLHNSYPAMPVPSLPTTIRTFYRCWPLCPDGRSSPVI GSPRYG (IN REF. 3).
SQ SEQUENCE 687 AA; 75672 MW; 24AFAD799DE0D396 CRC64;
Query Match 18.1%; Score 602.5; DB 1; Length 687;
Best Local Similarity 28.8%; Pred. No. 1.5e-34;
Matches 180; Conservative 131; Mismatches 220; Indels 95; Gaps 19;
OY 66 RQILKDVSLYVESGOIMCIIIGSSGSKTLLIDAMSGR--IGRAGTFIEGVYVNGALRRE 123
DB 110 KHLKNCVAVPBEILLAVWSSGAGKTTLLNALAFSPQISVSPSCRLNLCGPVAVK 169
OY 124 QFQCFSVIAGSDTLSSLVYRETLHYTALLAIRKNGSPQK--KTEAVAAELSLSHV 180
DB 170 EMQRCATVQODDLEFISLTAREHLIFQAM--VNRPHLYRQRRARADVOYIOELSLK 227
OY 181 ADRLIG-NYSLGIGSTGERRRVSTAOLADPKYMLFDEPTGDCMTANOIVLLVELA 239
DB 228 QHTIIGVPRKVGKSGSGERKRLAFASBLTDPPLICEDEPTSGDLSFAHVVQVKKLS 287
OY 240 RNRHIVLTTHOPSELEFOLFDTAILISFGEIIFCGPRAEMLPDNCQYCPREHSNPD 299
DB 288 QKGTIVLTTHOPSELEFELDKILLAMEGRVAFGLTSEAVDFSYGACPTNYNAD 347
OY 300 FYMDLTVSDVQSKEREIETSKRVOMISAYKKSACIKTKLNIEKMLKTLIPVPEFK 359
DB 348 FYVQVLAIV--VPREIESRDRIAKIDCNFAIS--KVADMQQLATKMLE--KPL 396
OY 360 DSP-----GVSKLGVLLRRYTRNLVNRKLAIVTRLLONLNGLEFL-LEFVLVR 410
DB 397 EQPENGYYTKATWPMQFRAVLMRSMLSYLKEPELVKVLIOGTWAILIGLIFGQQLTQ 456
OY 411 VLKGAIDRVGLTLOFGATPYTGMLNANVLEPVLRAVSDQESDGLYQKQMLAVLALH 470
DB 457 V---GVANINCAITLFLNMTFQVVFATINVFISELPFEMREASRLYRCOTYFLGK 513
OY 471 VLPESVATMIFSSVCWTGLQHEVARFGYFSALLAPHLIGELFVLGLIYONPNTIV 530
DB 514 ELPLFLVPLVETFAIAVPMIGLRAGVLFH-----FNCLATVTV--ANVS 556
OY 531 NSVVALSLAG-----VLVSGFLRNIOEMPFIKIIISYTFQKYSSEI 574
DB 557 TSEGYLLSCASSSTSMALSVGPYIIPLLFGLFGFLLNGSVYVLLKMLSYLSWFRV 616
OY 575 LVVNEFYGL---NFTGSSNVSVTNNMCAFTQIOIETKCP--GATSRRTANFLIIS 629
DB 617 LLIHQADVPEGEISCTSSNT-----TSPSSGVILLETINFSA--A 655
OY 630 FIP---ALVILGIVPKIRDHLIR 651
DB 656 DLPIDVGLAIL-IVSFRVLAYIALR 680
ID ABG1_HUMAN STANDARD; PRT; 678 AA.
AC P45844; Q9BXL0; Q9BXL1; Q9BXL3; Q9BXL4; Q9BXL6; Q9BXL7;
AC Q9BXL8; Q9BXL9;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 1 (White protein homolog)
DE (ATP-binding cassette transporter 8).
DE ABCG1 OR ABC8 OR WRT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE OF 3-678 FROM N.A. (ISOFORMS 1 AND 4).
RC TISSUE-Retina;
RX MEDLINE=96256850; PubMed=8659545;
RA Chen H.M., Rossier C., Lalliot M.D., Lynn A., Chakravarti A.,
RA Perrin G., Antonarakis S.E.;
RT "Cloning of the cDNA for a human homologue of the Drosophila white
RT gene and mapping to chromosome 21q22.3.";
RL Am. J. Hum. Genet. 59:66-75(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii R., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schilhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Saeki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordstok G., Hornischer K., Brandt P.,
RA Scharte M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ransner J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,
RA Weinmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehnach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=20408883; PubMed=10950923;
RA Berry A., Scott H.S., Kudoh J., Tallor I., Korostishevsky M.,
RA Wattenhofer M., Guipponi M., Barras C., Rossier C., Shibuya K.,
RA Wang J., Kawasaki K., Asakawa S., Minoshima S., Shimizu N.,
RA Antonarakis S.E., Bonne-Tamir B.;
RT "Refined localization of autosomal recessive nonsyndromic deafness
RT DFNB10 locus using 34 novel microsatellite markers, genomic
RT structure, and exclusion of six known genes in the region.";
RL Genomics 68:22-29(2000).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=21192304; PubMed=11279031;
RA Porsch-Oezcuerumez M., Langmann T., Heimerl S., Borsukova H.,
RA Kamlinski W.E., Drobnik W., Honer C., Schumacher C., Schmitz G.,
RT "The zinc finger protein 202 (ZNF202) is a transcriptional repressor
RT of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene
RT expression and a modulator of cellular lipid efflux.";
RL J. Biol. Chem. 276:12427-12433(2001).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS 2, 3, 4, 5, 6 AND 7).
RX MEDLINE=21092576; PubMed=11162488;
RA Lorkowski S., Rust S., Engel T., Jung E., Tegelkamp K., Galinski E.A.,
RA Assmann G., Cullen P.;
RT "Genomic sequence and structure of the human ABCG1 (ABCG8) gene.";
RL Biochem. Biophys. Res. Commun. 280:121-131(2001).
RN [6]
RP SEQUENCE OF 33-678 FROM N.A.
RC TISSUE-Fetal brain;
RX MEDLINE=97186700; PubMed=9034316;
RA Croop J.M., Tiller G.E., Fletcher J.A., Lux M.L., Raab E.,
RA Goldenson D., Arciniegas S., Son D., Wu R.;
RT "Isolation and characterization of a mammalian homolog of the

Db 248 DSASCFVSVSLMKGLAOGRSIICTIHOPSAKLELFDQVLVLSOGCCVYRGKVCNLVY 307
 QY 284 FNDGCPPEHSPNPFDFYMDLTSDYDTSKEREIETSKRVOMIESAYKSA-----I 334
 Db 308 LRDJLNCPTYNHNPADFEVAVASGEYQONSRLVAVNEGCDSDHKDLGDAEVPFL 367
 QY 335 CHKTUKNIERMKHLKTLPMVPKTPKDSFGV-----FSKLVLLRYVRLVNRKL 384
 Db 368 WHPRESEVKQYRKRLKGL-----RKDSSMEGCHSFSAKLTQCIIEFKRFLSLMRDSV 421
 QY 385 AVIRRLONLMGFLFELFVLVRSNVLKGAIDRVGILLYFVGATPTTGMLNANVLEPV 444
 Db 422 LTHLRITSHIGILGILLYGIGINEAKR--VLNSGFLFESMFLFMAALMPVLTPEPL 479
 QY 445 LRAVSDSDGLTOKNOMLAVLHVLPFSVAVPMIFSSVCYCTGLGHPVARGYSA 504
 Db 480 EMGVFLREHLNVMYSLAVYLAIAKTMADVPFQIMPPVACSTIVYMTSOPSDAVRFLFA 539
 QY 505 ALLAPHLIGFELTVLLGVQNPVIVNSVALLSIAGLVSGFLRNIOEMPIPEKTIISY 564
 Db 540 LGTWTSLVAOSIAGL-LIGAASTSLQVATFVGPTAIPVLLFSGFVSFDITPYLQNMXY 598
 QY 565 FTGPKYCEILVYVNEFGILNFTGSSNVSTTNPMCAFTGCIQFIETKCGCATSRFTMNF 624
 Db 599 ISYVRGFEVILS-ITGLD-----REDLHCDIDETCFH-OKSEALIRELDVENAKLYLDF 652
 QY 625 LILYSFIPALVILGIYV--FKIR 645
 Db 653 IVLGIFPISLRILVIFLVRKIR 675

RESULT 14
 ID WHIT_CERCA STANDARD; PRT; 679 AA.
 AC 017320:
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE White protein.
 GN W.
 OS Ceratitis capitata (Mediterranean fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Tephritidae; Tephritidae; Ceratitis.
 OC NCBI_TaxID=7213;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9613276; PubMed=8533095;
 RA Zyldebel L.J., Saccone G., Zacharopoulos A., Besansky N.J.,
 RA Favia G., Collins F.H., Louis C., Kafatos F.C.;
 RT "The white gene of Ceratitis capitata: a phenotypic marker for
 RT germ-line transformation".
 RL Science 270:2005-2007(1995).
 CC -1- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
 CC NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
 CC CELLS RESPONSIBLE FOR EYE COLOR.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: X89933; CAA61998.1;
 DR InterPro: IPR003439; ABC_tranport.
 DR InterPro: IPR005284; Pigment_permease.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD00006; ABC_tranport; 1.
 DR TIGRfams: TIGR00955; 3a01204; 1.

DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KW Pigment; ATP-binding; Transmembrane; Transport.
 FT NP_BIND 121 128 ATP (BY SIMILARITY).
 FT TRANSMEM 427 445 POTENTIAL.
 FT TRANSMEM 457 477 POTENTIAL.
 FT TRANSMEM 507 525 POTENTIAL.
 FT TRANSMEM 534 555 POTENTIAL.
 FT TRANSMEM 568 586 POTENTIAL.
 FT TRANSMEM 651 670 POTENTIAL.
 FT CARBOHYD 628 628 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 643 643 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 679 AA; 75145 MW; 3P9C8C78A835C40C CMC64;
 Query Match 17.8%; Score 591; DB 1; Length 679;
 Best Local Similarity 28.4%; Pred. No. 9.4e-34;
 Matches 176; Conservative 125; Mismatches 231; Indels 88; Gaps 18;
 QY 66 ROLKDVSLVYESGQIICIGSSGSGRTLLDAMSGRLGRAGTFLGEV--YVNGRALRE 123
 Db 101 KHLKNDGVAIPBELLAWMSSGAGKTLTLNLSAFSSKGVQISPTIRMLGHPVDAR 160
 QY 124 QPDCFSYVLOSDFLLSLVRETLYHTALAIRGNPGSFQK--KYEAVAEISLSTVA 181
 Db 161 EMQARCAVYQDDLEFISLAREHLIQAWRMFR-HHTQKQVARDQVQDLSLGRQ 219
 QY 182 DRLIG-NYSLGISTGERRRYSIAQLQDPKVLFDPTTGIDCMRTANOIVLVLELAR 240
 Db 220 NTLIGVGRYKGLSGGRKRLAFSEALTPEPLIDCEPISGLDSFMAHVVQVYKLKLSQ 279
 QY 241 RNRIVLTIHOPRSELPOLFDKTAIISFGLIGCGTFAEMDFPNDGCPPEHSPNPF 300
 Db 280 KGTIVLTIHOPSELELFDKTLMAEGRAVFGTGEAVDFSYIGATCPTVYTPADF 339
 QY 301 YMDLTSDYDTSKEREIETSKRVOMIESAYKSAICHRTLNIEERMKHLKTLPMVPKTRD 360
 Db 340 YVQVLAV---VPGREVSRRVAKICQNFVAGKYSREMDNFQ-----KLYKSGFKED 391
 QY 361 -----SPGVSKLGVLLRNVRLNVRNKLAVITRLLQNTIMGLFL-LFVVLVRSNVLK 413
 Db 392 ENETYKASWFMQFRAVLMSRWLSVLEKEPLVYKRLQTTMVAVLIGLIFLGQQLRYV-- 449
 QY 414 GATIDRGLLYXOPFAGAPRYGMLNANVLFVLAVYSQESQDGLYKRVOMLAVLVP 473
 Db 450 -GVANINGALFELFTNMTFQNSPATTIVTELEPFVFRERSRYRCDYFLGRTIHEL 508
 QY 474 FSVAVTMIFSSVCYWTGLHPEVARFEGYFAALLAPHLIGFELTVLLGVQNPVIVNSV 533
 Db 509 LFLVVPFLFAINAVPLGLRPGVDH-----FRLALVTVL--ANVSTSF 551
 QY 534 VALIS-----IAGVLVSGFLRNIOEMPIPEKTIISYFTGKYCEILVY 577
 Db 552 GYLISCASCSSTNALSVGPVILPFLFGLGFFLNSGVPYFKWLSTLSMFRYANDEGLI 611
 QY 578 NEFYGL---NFTGSSNVSTTNPMCAFTGCIQFIETKCG--GATSEFTANFL---ILYS 629
 Db 612 NQWADVAPGEITGLTSLT-----TGPSSEVILETINFASDLPFD. 652
 QY 630 FIP-ALVILGIYVFKIRHDL 648
 Db 653 FIGALLIIVG---FRISAVI 669
 RESULT 15
 ID ABG4_HUMAN STANDARD; PRT; 646 AA.
 AC 09H172:
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 4.
 GN ABG4 OR WHITE2.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21518231; PubMed=11606068;
 RA Engel T., Lorkowski S., Lueken A., Rust S., Schlueter B., Berger G.,
 RA Cullen P., Asmann G.;
 RT "The human ABCG4 gene is regulated by oxysterols and retinoids in
 RT monocyte-derived macrophages";
 RL Biochem. Biophys. Res. Commun. 288:483-488(2001);
 RN [2]
 RP SEQUENCE OF 20-646 FROM N.A.
 RC TISSUE=Dorsal root ganglion;
 RA Oldfield S., Lowry C.A., Lighman S.L.;
 RT "Cloning and expression of a novel mammalian white family
 RT ABC transporter: WHITE2";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: May be involved in macrophage lipid homeostasis.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ300465; CAC17140.1; -
 CC Gene; HGNC:13884; ABCG4.
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transporter.
 DR Pfam: PF00005; ABC_tran; 2.
 DR ProDom: PD000006; ABC_transporter; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 DR ATP-binding; Glycoprotein; Transmembrane; Transport.
 KW DOMAIN 1 393
 FT TRANSMEM 394 414
 FT DOMAIN 415 425
 FT TRANSMEM 426 446
 FT DOMAIN 447 472
 FT TRANSMEM 473 493
 FT DOMAIN 494 503
 FT TRANSMEM 504 524
 FT DOMAIN 525 532
 FT TRANSMEM 533 553
 FT DOMAIN 554 617
 FT TRANSMEM 618 638
 FT DOMAIN 639 646
 FT NP_BIND 102 109
 FT CARBOHYD 422 422
 FT SEQUENCE 646 AA; 71895 MW; 9CCEC6E150772611 CRC64;
 SQ
 Query Match 17.4%; Score 578.5; DB 1; Length 646;
 Best Local Similarity 27.1%; Pred. No. 6.6e-33;
 Matches 171; Conservative 126; Mismatches 274; Indels 59; Gaps 14;
 OY 33 PEPHSIGILHASTYSVSHRPMWMDITSCQOQTRQIINDKVSILYVESGQIMCITLSSGSGSK 92
 DB 54 PRRSAVDIEFVELSTSVREGPCW-----RKRGYKTLKCLSGKRCRRELIGIMGPSGAK 108
 OY 93 TTLIDAMSGRLGACFTLGEVYVNGRALREPODFSVLQSDTLSSLIVRETIHLYTA 152
 DB 109 STEFNIIAGY--RESGMMGQILYVNGRPRRLFRKMSCTIMQDDMLPHILYVLEAMVSA 166
 OY 153 LLAIRGNPGSFQKVEAV-----MAELISLVADRLIGNTSIGISTGERRVSTAAQ 206
 DB 167 NKLISEKO-----EVKKEIVTEITLALGLMSCSHRTAL-----LSGGQRRLAIALE 214
 OY 207 LLQDPKVMFLDEPTTGLDQMTANQIVLVLLVELARNRIVVLTIHQPRSELPQLEPKIAL 266

DB 215 LVNPPVMEFDEPTTSGIDSASCFQVVSIMKSLAOGGRITICTIHQPSAKLFEFEDLXYLI 274
 OY 267 SFCELIIFCGTAPAMLEFFNDGYPCEHNSPDPFYVDLTSVDFQSERELIETSKRYOMIE 326
 DB 275 SOGQCIFKGVNTLIYIKGLGICHTYINHPADFIIEVASG-----EGDLPMPLE 325
 OY 327 SAYKSAICHTKILKNIERMKHLKTLRPMVPEFKTQSD-----GVFSKIGVLLARVTRN 378
 DB 326 RAV-QNGICMAEKKSSPEKNEVPAPCPPEVVDLIESHTFATSTLQPCILFKRTPTS 384
 OY 379 LVNKLAVITRLQNLIMGLFLEFVLRKSNVLKCAIDRGLLYQVPGAPPYGMLAN 438
 DB 385 ILNDTVLTHLRFMSHVIGVLIGLYLHIGDDASK--VFNNMGCLFFSMILFALMLPT 442
 OY 439 VNLFPVLRVSDQSDGLYQKQMMALAYALHLPSVYATMFSSVCWTIGLHEVAR 498
 DB 443 VLFPLEMAVFMHEHLNWTSLKAYVLAQTMADVPQVVCPIVYMMTGQPAETSR 502
 OY 499 FGYSALLAPHLIGELTLVLLGIYQNPNIYNSVALLSIGVLVSGFLRNIOEMP 558
 DB 503 FILFSALATATALVAQSLG-LIGASNSLQVATFVGPYTAIPVLFSGFPVSFKTIPT 561
 OY 559 FKLIISFTYQKTCSEILVNEFYGL--NFTGSSNVSVTTPMCAFTGIGIETKTC 615
 DB 562 LQWSSYLSYVRVGEFEGVILT-IYGMERGDLTC-----LEERCPRFP-QSILRALDY 611
 OY 616 ATRSFPMNLLILYSFIPALVVIIGIVFKIR 645
 DB 612 EDKLYMDPLVLGIFPLALRLAYLVLR 641

Search completed: June 11, 2003, 09:12:37
 Job time : 18 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 09:11:47 ; Search time 18 Seconds
(without alignments)
1064.129 Million cell updates/sec

Title: US-09-989-981A-6

Perfect score: 3326

Sequence: 1 MGDSLTPGSGMGLQVNRG.....PALVILGIYVFKIRDLISR 651

Scoring table: BLOSUM62

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: 1 Issued Patents, AA:*
2: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/CTUS.COMB.pep:*
7: /cgn2_6/ptodata/1/1aa/Dackfilles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	682.5	20.5	655	4	US-09-245-808-1
2	244	7.3	248	4	US-09-134-001C-3731
3	244	7.3	1280	2	US-08-752-447-2
4	244	7.3	1280	4	US-09-316-167-2
5	240	7.2	1279	2	US-08-784-649A-2
6	240	7.2	1280	2	US-08-583-276-19
7	240	7.2	1280	6	5206352-4
8	239.5	7.2	1684	3	US-08-665-259-25
9	239.5	7.2	1684	3	US-08-762-500-25
10	239.5	7.2	1704	3	US-08-762-500-75
11	230.5	6.9	242	4	US-09-134-001C-3832
12	229	6.9	460	4	US-09-134-001C-3369
13	229	6.9	1307	1	US-08-395-246C-2
14	220.5	6.6	273	4	US-09-134-001C-3641
15	220	6.6	1457	3	US-08-665-259-27
16	220	6.6	1457	3	US-08-665-500-27
17	218	6.6	1275	3	US-09-120-513-2
18	218	6.6	1275	4	US-09-450-105-2
19	213	6.4	228	4	US-09-134-001C-3373
20	211	6.3	270	4	US-09-134-001C-5561
21	210	6.3	382	4	US-09-134-001C-3439
22	209.5	6.3	244	4	US-08-919-573-2
23	209.5	6.3	244	4	US-08-919-573-4
24	209.5	6.3	341	4	US-09-134-001C-4042
25	208.5	6.3	202	4	US-08-858-207A-332
26	208.5	6.3	229	4	US-09-134-001C-4113
27	207.5	6.2	246	4	US-09-627-376-15

28	206.5	6.2	233	4	US-08-993-825-2	Sequence 2, Appl1
29	206.5	6.2	1476	4	US-09-256-703-2	Sequence 2, Appl1
30	206.5	6.2	1480	1	US-07-637-621-2	Sequence 2, Appl1
31	206.5	6.2	1480	2	US-08-951-912-2	Sequence 2, Appl1
32	206.5	6.2	1480	4	US-08-681-838A-2	Sequence 2, Appl1
33	206.5	6.2	1480	4	US-08-681-838A-3	Sequence 3, Appl1
34	206.5	6.2	1480	4	US-09-174-077-2	Sequence 2, Appl1
35	206.5	6.2	1480	6	5240846-5	Patent No. 5240846
36	206	6.2	406	4	US-09-134-001C-3570	Sequence 3570, Ap
37	206	6.2	1479	2	US-08-951-912-4	Sequence 4, Appl1
38	206	6.2	1479	4	US-09-174-077-4	Sequence 4, Appl1
39	205.5	6.2	1480	4	US-09-425-453A-6	Sequence 6, Appl1
40	205.5	6.2	1480	4	US-09-425-453A-12	Sequence 12, Appl1
41	205.5	6.2	1480	4	US-09-425-453A-20	Sequence 20, Appl1
42	205	6.2	225	4	US-09-134-001C-3215	Sequence 3215, Ap
43	204.5	6.1	1308	2	US-08-996-644-2	Sequence 2, Appl1
44	204.5	6.1	1308	3	US-09-352-552-2	Sequence 2, Appl1
45	204.5	6.1	1480	1	US-08-136-742A-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1	US-09-245-808-1	Sequence 1, Application US/09245808
Patent No. 6313277		
GENERAL INFORMATION:		
APPLICANT: Doyle, L. Austin		
APPLICANT: Adruzzo, Lynne V.		
APPLICANT: Ross, Douglas D.		
TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which		
FILE REFERENCE: Ross DMD conversion		
CURRENT APPLICATION NUMBER: US/09/245,808		
CURRENT FILING DATE: 1999-02-05		
EARLIER APPLICATION NUMBER: 60/073763		
EARLIER FILING DATE: 1998-02-05		
NUMBER OF SEQ ID NOS: 7		
SOFTWARE: Patentln Ver. 2.0		
SEQ ID NO 1		
LENGTH: 655		
TYPE: PRT		
ORGANISM: Human MCF-7/AdrVP cells		
US-09-245-808-1		
Query Match	20.5%; Score 682.5; DB 4; Length 655;	
Best local similarity	29.2%; Pred. No. 1.9e-64;	
Matches	182; Conservative 138; Mismatches 249; Indels 55; Gaps 18;	
OY	21 SSSLEGAPATAP---EPHSGLILNHSYSVSHRPMDITSCROOQROIIDKDYLYVE 77	
DB	13 SGGNNGFPATASNDLKAFTGCAVLSFHNICRYVTKSGFLCRPVEKEILSNIGIMK 72	
OY	78 SCQIHCIIASSSGGTTLLDANSGLRGAGTFLGEVYVNGALRBDQFDCSYVQSDT 137	
DB	73 PG-LNAILDPTGGGSSCLDYLAAKRDSCG-LSGCVLLING-APRANFCNGGYVQDDV 129	
OY	138 LLSLTVRETLVLTLLAIRGNPG-SPOKKEAVMAELSLSHVADRILGNSTLGSISTG 196	
DB	130 VAGTITVENLQFSALRIATMTWHNEKRNIRVIOELGIDKLVKSGQFINGVSG 189	
OY	197 ERRRVISAQLQDPKVMFLDEPTGLDPMANOIVLVLLVELARRNRIVLTIHOPRSL 256	
DB	190 EKKRISIMELTDSIIFDEPTGLDOSTANAVLLIKRMSKGRITISIHPRISI 249	
OY	257 FQLPKVAILSFGLIFGCTPAEMLDFFNDGCGPCPEHSNPDFYMDLTSDVTO---SK 312	
DB	250 FKLFDSTLTLASGRIMFGPAQOEALGYFESAGYHCEAVNNPADFLDIINGDSTAVLNR 309	
OY	313 ERE-----IESKR-----VQIESAYKKSALCHT-----LKNIEKMKHKTLPMPVF 356	
DB	310 EEDFKATEITIEPSKODKPLIEKLAETIYVNSSYFKETKALQLSGEKKRKITVKEISY 369	

357 KTKDSGVYFKLGLRLRVNKLAVITLRLNGLFT--LEFVLRVSNVLKG 414
370 TT-----SFCOLRWKSRKSKFKNLGNPOASIAQIIVTVVLGIVGAIYFGKJNST----- 421
415 AIODRVGLLYQVFGATPTGMLNANVLEFVLKAVSDOSOGILQKQKQMLAYAL-HVLP 473
422 GIQNAQVLE-FLTTCOFSSVSAVLEFVEKLEFHEHYISGYRVSSYFLGKLKLDLPL 480
474 FSVVATNIFSSVCYWTGLLHEVAVRFGFSAALLAPHLIGELTLLGIQONPYNVSV 533
481 WMLPSTIIFCTIYVFNGLKRAKADAFEMFTLM---WVAVSASMALAIAGOSVSV 537
534 VALLISAGV--LVGSGFLRNTQEMPIPKIISYFFQYCEIIVNVEFYGLNFTCGSSN 591
538 TLMATICEFVFMVIFSGILVNLTTIASMLSMLOYSIPRYGTALQHNELQNCQPG--- 594
592 VSVTTPMCAFTQGIQIFLEKTCPG 615
595 LMATGNPCNTA-----TCTG 610

RESULT 2

US-09-134-001C-3731
Sequence 3731, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134, 001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3731
LENGTH: 248
TYPE: PR
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3731

Query Match 7.3%; Score 244; DB 4; Length 248;
Best Local Similarity 29.3%; Pred. No. 8,4e-18;
Matches 70; Conservative 53; Mismatches 92; Indels 24; Gaps 9;

67 QILKDVSLYVSGQIMCIISSSGSKTLLDMSG-RLGRAGTFLGEVYVNGRRLRE-- 123
20 EVIKGIDLKINOGVAVLLIGRSGSKTLLMINALEIPEGT---VYVNGMTYNTKDK 75
124 ----QFODCFSYVLOSDTLLSLVRETLHTALLATRBGNPSFOKKVEVMAELSLSH 179
76 KSQLKVAQOQSGMYQNTNLPFKSALENV-MEGILTVKMKKATANEAMNLAKVGIVH 134
180 VADRLIGNYSLSIGISTERRRVSIAQLDPPKYMALDEPTTGIDCTANQIVLVLELA 239
135 VKDO--RPHALSG---GQOQRVAIARALANPKMALDEPTSAIDPELVNDYLVKIELA 189
240 RRRNIVYLTIHQPSSELPFLQDKALISFEGELIFCGPAEMLEDFNDCGIPCEHSNPF 298
190 DEGMTAVIVTHEMFAK-EVSNQIAFTHEGVIAEGQTPG---DIFN---HPKTELQRF 241

RESULT 3

US-08-752-447-2
Sequence 2, Application US/08752447
Patent No. 5994088
GENERAL INFORMATION:
APPLICANT: Mechettner, Eugene
APPLICANT: Roninson, Igor B
TITLE OF INVENTION: Methods and Reagents for Preparing and

TITLE OF INVENTION: Using Immunological Agents Specific for P-glycoprotein

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSER: McDowell Boehnen Hubert & Berghoff Ltd.

STREET: 300 South Wacker Drive, Seventh Floor

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/752,447

FILING DATE: 15-NOV-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: No. 5994088nan, Kevin E

REGISTRATION NUMBER: 35,303

REFERENCE/DOCKET NUMBER: 95,1121

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-913-9808

TELEFAX: 312-913-9808

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1280 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-752-447-2
Query Match 7.3%; Score 244; DB 2; Length 1280;
Best Local Similarity 20.7%; Pred. No. 1.5e-16;
Matches 153; Conservative 106; Mismatches 229; Indels 250; Gaps 32;

41 LHASYSYSHRYRPMWMDITSCNQWTRQILKDVSLYVSGQIMCIISSSGSKTLLDAMS 100
397 VHFSPSRKRYK-----ILKGLNLTQSGGVAVLVGSGGSKSTVOILQ 441
101 GRIGRAGTFLGEVYVNGRRLR-REOPDQCFSYVLOSDTLLSLVRETLHTYALLAIR 153
442 -RL--YDPTGCMVSDQDRTIRTVRLRELIGVSGEPVLFATTIENIRY----- 490
159 GNPSPFKRYEAVNAE-----LSLHVADRLIGNYSLGISTGERRRVSIAAQLDOP 211
491 GRENVTDEIEKAVKANAYDFIMKLPKKEDTLVGERG-AQLSGGOKORIAARALVRNP 549
212 KVMLEDEPTGIDCMQNOIVLVLELARNRIVYLTIHQPSSELPFLQDKALISLSEL 271
550 KIILLDEATSLD-TESEAVYQVALDKARKGRTTIVAH--REATVNRADVIAGFDGVI 606
272 IFCGTPAEMLEDFNDCGYPCEHSNPFYMDLTSVDQSKEREIETSK----- 320
607 VERGNHDELM-----KEKGITFKLYTMTQACNEVELNAADESKSEIDALE 652
321 -----RQIMISAKKSAIKCTIKNIERNKHLTLPPVFP-----KTKDSPG 363
653 MSSNDSRSSLRKRSTRSVRSQOARHKLSTREALD--ESIPVSFWIRIMKILNLTMPY 710
364 -----VFSK-LGVLR-----RVTRN-----LYRNKLAVITR 389
711 FVYGVFCAILINGGLQPAFAIIFSKIIIVFPRIDDEPKRONSLSLFLALGIITFT 770
390 LQNLING-----LFLFLVYLRVRSNV-----LNGAIO 417
771 FLOGFTFGKAGEILTKRLRYVFRSMRGVSNFHDCKNTGALTRLANDAQVYGAIG 830
418 DRV-----GLLYQVFGATPTGMLNANVLEFVL-----RAYSQOE 452
831 SRLAVITONTANIGTGIITISFTYGMOTLTLLAI--VPITAIAGVEMKFAQALDKR 888

QY 453 SODGL-----YOKWOMLAVAHV-----LPSVATM 480
DB 889 ELEAGKATATEINERTVSLTQOFKEHMYAOSLQVPRNSLRKAHIFGITTSFTQAM 948
QY 481 IFSSV--CYWTLGLHPEVARFGYSAALLAPHLIGER--LTLVLGLIYQONNIYNSV--- 533
DB 949 MYFSYAGCF-----RFG---ATLVAKHKLMSFEDVLVFSAVFGAMAVGOYSSF 994
QY 534 -----VALLSIAGVLVSGFLRNIOEMPIPKIISYTFQKYSFELVNEFGYGLNFTCG 588
DB 995 APYAKAKISAHHIM-----IIEKPTL-----IDSYSTEGMLPNTLEG-NVTEG 1038
QY 589 SSNVSTTNPMCAFTGCI 606
DB 1039 EVENNYPTRPDIPLYQGL 1056

RESULT 4
US-09-316-167-2
Sequence 2, Application US/09316167
Patent No. 6365357
GENERAL INFORMATION:
APPLICANT: Mechelner, Eugene
APPLICANT: Roninson, Igor B
TITLE OF INVENTION: Methods and Reagents for Preparing and
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESSES:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff Ltd.
STREET: 300 South Wacker Drive, Seventh Floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/316,167
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/752,447
FILING DATE: 15-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: No. 6365357nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 95,1121
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-9808
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1280 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-316-167-2

Query Match 7.3%; Score 244; DB 4; Length 1280;
Best local Similarity 20.7%; Pred. No. 1.5e-16;
Matches 153; Conservative 106; Mismatches 229; Indels 250; Gaps 32;

QY 41 LHASYSVSHRVRPMWDITSCROQWTRQILKDVSLYEGQIMCILLSSGSGKTLTDAMS 100
DB 397 VHSYSPSRKEVK-----ILKGLMLKQSGQTVALLVNSGSGKSTTVQIMQ 441
QY 101 GRGRAGTFLGEYTVNGRALR--REOPDCESTYVLQSDTLSSLVARETLHYTALLAIR 158
DB 442 -RL--YDPTGEMVSDGQDIFRTINVRFLREITIGVSGQEPVLFATTAENIRY----- 490

QY 159 GNPGSFQKVEANVAE-----LSLSVADRILGNSTSGISGERRRSIAAQLQDP 211
DB 491 GRENVTMEDEKAEKAEANAYPIIMKLPKFTLVGGERG-AQLSGQORAIADALRNRP 549
QY 212 KVMLEDEPTTGIDCTANQIVLVLELARNRIVLTIHOPRSELFQFDKIALISFGL 271
DB 550 KILLIDEATSLD--TESAVQVVALDKARKRTTIVIAH--REFATRVADYIAFGDCVI 606
QY 272 IFCTPAPMLDFPNDCCGIPCEPHSNPEDEYDLTSVDTQSKERELETISK----- 320
DB 607 VEKGNHDELM-----KKGITYEKLVTMOTAGNEVELENADESKSEIDALE 652
QY 321 -----RYQMIESAYKKSALCHETLKNIERMKLTLNVPF-----KTKSPG 363
DB 653 MSSNDSNSSLIRKSTRSRVSGOAHKRLSTKRALD--ESIPVSPWRIMKLNLTMPY 710
QY 364 -----VESK-LGVLR-----RVTRN-----LVNKLAVTR 389
DB 711 FVGVFCAIINGGLQPNALIFSKITIGVTRIDDEPTRYKNSNLFSLFLALGITSFTF 770
QY 390 LLQNLIMG-----LELLEFVLNRSNV-----LKGAIQ 417
DB 771 FLOGFTFGKAGEILTKRLRYVWFRSMLEQDVSFHDPKNTGALTTRLANDAAYKGAIG 830
QY 418 DRV-----GLLYQVGAIPYTGMLNAYNLPVL-----RAVSDQ 452
DB 831 SRLAVTQNMANTGTGIIISFIYGMQLTLILLAT--VPIAIAGVEMKEMFACQALDKK 888
QY 453 SODGL-----YOKWOMLAVAHV-----LPSVATM 480
DB 889 ELEAGKATATEINERTVSLTQOFKEHMYAOSLQVPRNSLRKAHIFGITTSFTQAM 948
QY 481 IFSSV--CYWTLGLHPEVARFGYSAALLAPHLIGER--LTLVLGLIYQONNIYNSV--- 533
DB 949 MYFSYAGCF-----RFG---ATLVAKHKLMSFEDVLVFSAVFGAMAVGOYSSF 994
QY 534 -----VALLSIAGVLVSGFLRNIOEMPIPKIISYTFQKYSFELVNEFGYGLNFTCG 588
DB 995 APYAKAKISAHHIM-----IIEKPTL-----IDSYSTEGMLPNTLEG-NVTEG 1038
QY 589 SSNVSTTNPMCAFTGCI 606
DB 1039 EVENNYPTRPDIPLYQGL 1056

RESULT 5
US-08-784-649A-2
Sequence 2, Application US/08784649A
Patent No. 5830697
GENERAL INFORMATION:
APPLICANT: SIKIC, Branislav I
APPLICANT: Chen, Gang
TITLE OF INVENTION: P-GLYCOPROTEIN MUTANT RESISTANT TO
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Fish & Richardson
STREET: 2200 Sand Hill Road
CITY: Menlo Park
STATE: CA
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/784,649A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J

QY 392 ONLIMG-----LFLFEVLVRNSV-----LKGAIODR 419
 Db 773 OGFEFGAGAGELTFRRLRYMFRSMRLRODVSWFDDPKNTGALTTRLANDAQAQVGAIGSR 832
 QY 420 V-----GLXYQVGAATPYTGLNAVNLPVL-----RAVSDESQ 454
 Db 833 LAVITONIANIGTGIIISFYGMQLTLLALAI--VPILAIAGVEMKMLSGQALKDKKEL 890
 QY 455 DGL-----YOKOMMLATYALH-----LPSVYATWIF 482
 Db 891 EGAGKATEAEIENFRIVSLTOBQEFHMTAOSIQVPEYRSLRKAHIFGITFSTFOAMY 950
 QY 483 SSV--CYWTGLHPEVARFGYFSALLAPHLIGEF--LTLVLGIVONPNIVNSV-----533
 Db 951 FSVAGCF-----RFG-----AYLVAKHMSFEDEVLLVFSAVFGAAGVGVSSFP 996
 QY 534 ---VALLISGVLVSGFLRNIOEMPPIPKIISYTFQKCYSEILVNEFYGLNFTGSS 590
 Db 997 DYAKAKISAHHIIM-----IIKTPPL-----IDYSTBGLMPTNLEG--NVTGGEV 1040
 QY 591 NVSVTTPKCAFTQGI 606
 Db 1041 VFNYTPRPDIPLVLOGL 1056

RESULT 7

5206352-4
 Patent No. 5206352

Applicant: Konlinson, Igor B.; Pastan, Ira H.; Gottesman, Michael M.

SEQUENCES ASSOCIATED WITH MULTIDRUG RESISTANCE IN HUMAN CELLS
 NUMBER OF SEQUENCES: 4

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/622,836

FILING DATE: 24-SEP-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 892,575

FILING DATE: 01-AUG-1986

APPLICATION NUMBER: 845,610

FILING DATE: 28-MAR-1986

SEQ ID NO: 4

LENGTH: 1280

5206352-4

Query Match

Best Local Similarity 7.28; Score 240; DB 6; Length 1280;

Matches 152; Conservative 108; Mismatches 230; Indels 246; Gaps 32;

QY 41 LHASYSVSHRVRMMWITSCROQWTRQIILKDVSLVYESGQIMCIISSGSGKTTLDAMS 100
 Db 397 VHSYSRKREVK-----ILKGLMKVQSGQVALVNGSGCKSTTVQMQ 441
 QY 101 GRLGAGTFLGEVYVNGRLR--REOFODFSYVLASDTLSSLVRETLHTYALALAIR 158
 Db 442 -RL--YDPEGMVSVGODIRTNVFLREIIGVSOEPLFTTAENIRY-----490
 QY 159 GNGSFOKKEAVMAE-----LSLSHADRLIGNSLIGISGERRRVSIAAQLLODP 211
 Db 491 GRENVTMDEIKAVKEANAYDFIMKLPHEFDLVGBRG--AQLSGGOKORIAALVALVRN 549
 QY 212 KVALPDEPTTGLDCMTANQIYVLLVRLARNRIVATITQPRSELQOLDKAIISFGL 271
 Db 550 KILLDEATSAID--TESAVVOVALDKARKGRTTIVIAH--RLSTVRNADVIAGFDGVI 606
 QY 272 IFCGPAEMADFEFNDGYPCEPHSNPFDFYMDLTSVDYOSKEEIEITS-----KRVQ 323
 Db 607 VEGKNDELH-----KKGITYFLVYQTAGNEVELENAADESGSEIDALE 652
 QY 324 MIESAYKKAICHKTKIKNTERMKH-----KTLPMVPF-----KTKDSPG--363
 Db 653 MSSNDSRSSLIIRKSTRSRVSGQAODRKISTFEALDESIPVPSFRIKMLNTEMPYEV 712

QY 364 -----VFESK--LGVLLR-----RVTRN-----LVNKKIATVITRLL 391
 Db 713 VGFCAIITNGLOPAPAIITFSKITGVTRIDDPETRKONSLFSLFLAIGIISFYTFPL 772
 QY 392 ONLIMG-----LFLFEVLVRNSV-----LKGAIODR 419
 Db 773 OGFEFGAGAGELTFRRLRYMFRSMRLRODVSWFDDPKNTGALTTRLANDAQAQVGAIGSR 832
 QY 420 V-----GLXYQVGAATPYTGLNAVNLPVL-----RAVSDESQ 454
 Db 833 LAVITONIANIGTGIIISFYGMQLTLLALAI--VPILAIAGVEMKMLSGQALKDKKEL 890
 QY 455 DGL-----YOKOMMLATYALH-----LPSVYATWIF 482
 Db 891 EGAGKATEAEIENFRIVSLTOBQEFHMTAOSIQVPEYRSLRKAHIFGITFSTFOAMY 950
 QY 483 SSV--CYWTGLHPEVARFGYFSALLAPHLIGEF--LTLVLGIVONPNIVNSV-----533
 Db 951 FSVAGCF-----RFG-----AYLVAKHMSFEDEVLLVFSAVFGAAGVGVSSFP 996
 QY 534 ---VALLISGVLVSGFLRNIOEMPPIPKIISYTFQKCYSEILVNEFYGLNFTGSS 590
 Db 997 DYAKAKISAHHIIM-----IIKTPPL-----IDYSTBGLMPTNLEG--NVTGGEV 1040
 QY 591 NVSVTTPKCAFTQGI 606
 Db 1041 VFNYTPRPDIPLVLOGL 1056

RESULT 8

US-08-665-259-25

Sequence 25, Application US/08665259

Patent No. 6028173

GENERAL INFORMATION:

APPLICANT: Landes, Gregory M.

APPLICANT: Burn, Timothy C.

APPLICANT: Connors, Timothy D.

APPLICANT: Dackowski, William R.

APPLICANT: Van Raay, Terence J.

APPLICANT: Kilinger, Katherine W.

TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES.

NUMBER OF SEQUENCES: 73

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENZYME CORPORATION

STREET: One Mountain Road

CITY: Framingham

STATE: Massachusetts

COUNTRY: United States of America

ZIP: 01701

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/665,259

FILING DATE: 17-JUN-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Dugan, Deborah A.

REGISTRATION NUMBER: 37,315

REFERENCE/DOCKET NUMBER: IGS-9.1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (508) 872-8400

TELEFAX: (508) 872-5415

INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:

LENGTH: 1684 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

Wed Jun 11 09:51:14 2003

us-09-989-981a-6.rai

Page 7

Query Match	7.28;	Score 239.5;	DB 3;	Length 1704;
Best Local Similarity	32.58;	Pred. No. 7.5e-16;		
Matches 78; Conservative	43;	Mismatches 100;	Indels 19;	Gaps 9

[illegible]

RESULT 11

```

US-09-134-001C-3832
: Sequence 3832, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 3832
: LENGTH: 242
: TYPE: PRF
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3832

```

Query Match	6.9%	Score 230.5;	DB 4;	Length 242;
Best Local Similarity	26.1%;	Pred. NO. 2.3e-16;		
Matches 64;	Conservative 56;	Mismatches 90;	Indels 35;	Gaps 7

[illegible]

RESULT 12

US-09-134-001C-3369
; Sequence 3369, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:

```

: APPLICANT: Lynn Doucellette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 3369
: LENGTH: 460
: TYPE: PRN
: ORGANISM: Staphylococcus epidermidis
: US-09-134-001C-3369

```

Query Match	6.9%:	Score 229;	DB 4;	Length 460;
Best Local Similarity	25.2%:	Pred. No. 1e-15;		
Matches 77;	Conservative	66;	Mismatches 121;	Indels 42;
				Gaps 10

```

QY      66  ROLIKVSLIYSSGGOIMCILTSSSGCKTLLDASGRGRGTFLEVVYNGALRR--- 122
Db      57  KRAVDNISLIDIOGEEIAFICTSGSGKTTAL-RMINRIETAWD--GOIMNKGDVRRMNP 113
QY      123  EGFODEFSYVLOSDFTLSSLVFWEHLHTALALRBMPSGFSFOKVAAVAEISLHVAD 182
Db      114  VELRSIGIVYIOQIGLPMHTIREN-----YLVPKLLAKSKKEKDEKAKELIKLVDPFE 168
QY      183  RLIGNSLIGISTGERRRYSIAAQLLODPKVMLEFDEPTTGIDCKTANQIVLVELARRN 242
Db      169  EYLDPRP-AELSGGQQRIGVYRBALAAQDIIIMDEPRGALDPIYTRBOTDYKELQOQL 227
QY      243  RIYVLTIHQNSELEFQLEPDKTALISFEGELFCGTPAEMLKDFPNDGCGYCPCHSNPDPFYM 302
Db      228  GKTEFIIVTDHMEAKLADKICIMCSIKGVVOYDPRDNL-----RYRAN-----DFVR 275
QY      303  DLSVDYQSKERELEIETSKRVOMIESAYKKSALCKRTLK-----NIERMKHLKTLPM 353
Db      276  DFL-----GQNLIDLRPNMKSVEBSAMIKPV---TVKADSLNDVAVINMKTTRAVDTIFV 326
QY      354  VEPKTK 359
Db      327  VANNQNK 332

```

RESULT 13

US-08-395-246C-2
Sequence 2, Application US/08395246C
Patent No. 5773214
GENERAL INFORMATION:
APPLICANT: Peery, Robert B.
APPLICANT: Skatrud, Paul L.
TITLE OF INVENTION: MULTIPLE DRUG RESISTANCE GENE OF
TITLE OF INVENTION: ASPERGILLUS FLAVUS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: IN
COUNTRY: USA,
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/395,246C
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

```

NAME: Plant, Thomas G.
REGISTRATION NUMBER: 35784
REFERENCE/DOCKET NUMBER: X9683
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
TELEFAX: 317-277-1917
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1307 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-395-246C-2

Query Match
Best local Similarity 23.0%; Score 229; DB 1; Length 1307;
Matches 127; Conservative 110; Mismatches 188; Indels 126; Gaps 29;

QY 68 ILKDVSLVYESGQIMCIISSGSGKTTLLDMSGRIGR-AGFFLEGVY-----VNGRALRR 122
DB 407 VAHDLSCYIPAGKTTAFVGPSSGSKSTIISLERFYDPVAGTIMGHDIQTLNLRMLRQ 466
QY 123 EFOFDFSYVLQSDTLSSLYRETLHYTALLAIRGNPSPFO--KVEAV--MAE-----174
DB 467 Q-----MSLYSQEPRLFAT-TIAENIRY-GIIGSRFEKSTYEIRKRVPAARMAAHDF 519
QY 175 -LSLSHVADRLIGNSLGISTGERRRVSIAAQLQDPKVMLEFDEPTTGDCMTANQIV 233
DB 520 IMALNGYDPTNIESLSG---GOKORIMARAIIMDPKILLDETSALD-TKSEKLVQ 575
QY 234 LLVELARRRRIYVLTIHQRSELFQDKIALISFGEILFGSTPAEMLDFFNDGCPCE 293
DB 576 AALDKASKGRTTIVIAH--RLSTIQKANYIIVLANGQIYEQCPHEILMD-----622
QY 294 HSNPDFYMDLTSVTSQSKEREIETSRQMTESAYKKSALCHTKLTKNERKHLKILPM 353
DB 623 ---RGIYCDMEAH-EIKKRSRKRSKRSQ-----LTNISP-----KHNP 661
QY 354 VPFKTKDSFG-----VF-----SKLGVLRRVTRNLYRNKLAIVTRLLQMLINGLFTL 401
DB 662 TFFFKDYFGDDESDIYSILSDASDIGHTGEKQRPYSRMSLSHLMQPKVEAYSFWTL 721
QY 402 -----FVLVRNSVNLKCATODRGILIXOPGATPYTGMLNANVLEFVLRV 448
DB 722 KFLASFRNPEKPFLLIGLCASTLAGIOPSOAVLFAANVSTLSLPLE---YPKLR--775
QY 449 SDQESQDGLYQKQMLAYALVLPVAVTMI-FSS--VCY-----WTGLHPEVARE 499
DB 776 ---HDANFQMLFPLMIGIVSLVSVQGTLEFASSEKMYRANSQAFVILHODISFF 830
QY 500 -----GYFSALIA--PHLIGELTIVLLG--IVONPNIVNSV-----VALLSI 539
DB 831 DQOENTTCALTRATLSAGTKRELNG--ISGVTLGTILIVSNLVASIGVALVIGMKLALVCI 888
QY 540 AGV--LVGSGFLR 550
DB 889 SAVPALMCGFVR 901

RESULT 14
US-09-134-001C-3641
Sequence 3641, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779

```

```

PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3641
LENGTH: 273
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3641

Query Match
Best local Similarity 24.7%; Score 220.5; DB 4; Length 273;
Matches 62; Conservative 52; Mismatches 82; Indels 55; Gaps 7;

QY 66 ROLIKDVSLVYESGQIMCIISSGSGKTTLLDMSGRIGRAGTFLGEVYVNGRALRREP 125
DB 31 KHLLENINIKINGEELAIYPRGACKSTILKVI---LGLLPQKEIIVDGPFGKNS 87
QY 126 QDCFSYVLQ-----SDTLSSLYRETLHYTALLAIRGNPSPFOKVEAVM 172
DB 88 SLKTSYVOKASAFNAGFPASVKEVYLSGLTKKKL-----FQRFNNNDQ-KYTKVL 139
QY 173 AELTSHVADRLIGNSLGISTGERRRVSIAAQLQDPKVMLEFDEPTTGDCMTANQIV 232
DB 140 ERLNLSHLT-----NKNIELSGGQOQRLIARALISDPSVYLDEPTNGIDAKHVSKEY 194
QY 233 VLLVELARRRRIYVLTIH-----QPRSELFQDLKIALISFGEIL 271
DB 195 DPLDKAKKGCITIIIVTHDIGVADTATEVACLNKHLHFGSTEARKSIDEVEISK-----250
QY 272 IFCGTPAEMLD 282
DB 251 -IYGHPIQFVD 260

RESULT 15
US-08-665-259-27
Sequence 27, Application US/08665259
Patent No. 6028173
GENERAL INFORMATION:
APPLICANT: Landes, Gregory M.
APPLICANT: Burn, Timothy C.
APPLICANT: Connors, Timothy D.
APPLICANT: Dackowski, William R.
APPLICANT: Van Raay, Terence J.
APPLICANT: Klinger, Katherine W.
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSER: GENZYME CORPORATION
STREET: One Mountain Road
CITY: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665,259
CLASSIFICATION: 435
FILING DATE: 17-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Dugan, Deborah A.
REGISTRATION NUMBER: 37,315
REFERENCE/DOCKET NUMBER: IGS-9.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 1457 amino acids

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 09:08:12 ; Search time 39 Seconds

(Without alignments)
3439.402 Million cell updates/sec

Title: us-09-989-981a-6

Perfect score: 3326

Sequence: 1 MGDLSSLPFGSGKGLQVNRG.....PALVILGIYVFKIRDLISR 651

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: 1: SP_ARCHAEA:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	704	21.2	673	11 OBR543	08t543 mus musculu
2	680.5	20.5	655	4 O96TAR	096tar8 homo sapien
3	672.5	20.2	655	4 O96LDB	096ld6 homo sapien
4	662	19.9	801	5 O8T691	08t691 dictyostell
5	660	19.8	657	11 O9R004	09r004 mus musculu
6	649.5	19.5	725	10 O9M3D6	09m3d6 arabidopsis
7	628	18.9	691	10 O8RW19	08rw19 arabidopsis
8	622	18.7	668	10 O8AR04	08ar04 oryza sativ
9	618.5	18.6	672	10 O9L182	09l182 arabidopsis
10	617	18.6	727	10 O9FNBS	09fnbs arabidopsis
11	615	18.5	692	5 P91892	09f182 arabidopsis
12	614.5	18.5	703	10 O8RXNO	08rxno aedes aegypt
13	614	18.5	594	10 O9LJG3	09ljg3 arabidopsis
14	614	18.5	720	10 O9M2V7	09m2v7 arabidopsis
15	610.5	18.4	725	10 O9ZU35	09zu35 arabidopsis
16	610.5	18.4	725	10 O9ASR9	09asr9 arabidopsis

17	610	18.3	679	5 O9BH97	09bh97 ceratitis c
18	608	18.3	708	10 O9M2V5	09m2v5 arabidopsis
19	602.5	18.1	654	10 O9L1W2	09l1w2 oryza sativ
20	600.5	18.1	670	5 O77423	077423 bacterocera
21	600	18.0	602	5 O9VC15	09vc15 dirosophila
22	597	17.9	590	10 O9MAH4	09mah4 arabidopsis
23	595.5	17.9	658	5 O16574	016574 caenorhabdi
24	595.5	17.9	687	5 O94360	094360 dirosophila
25	595.5	17.9	785	4 O96L76	096l76 homo sapien
26	592	17.8	610	5 P90746	090746 caenorhabdi
27	591.5	17.8	740	10 O8O946	08o946 arabidopsis
28	589.5	17.7	646	10 O9C6R7	09c6r7 arabidopsis
29	588.5	17.7	648	10 O9C6W5	09c6w5 arabidopsis
30	587.5	17.7	627	11 O9LWA9	09lwa9 mus musculu
31	587.5	17.7	646	11 O9VBS9	09vbs9 mus musculu
32	584	17.6	755	10 O9ZUT0	09zut0 arabidopsis
33	583.5	17.5	687	10 O9C8K2	09c8k2 arabidopsis
34	581.5	17.5	687	5 O9NH94	09nh94 bombyx mori
35	580.5	17.5	669	5 O8WRP2	08wrp2 tribolium c
36	578	17.4	705	10 O9L1W1	09l1w1 oryza sativ
37	575.5	17.3	649	10 O9S1T6	09s1t6 arabidopsis
38	575.5	17.3	666	11 O9EPG9	09epg9 rattus norv
39	573.5	17.2	669	5 O8WR11	08wr11 tribolium c
40	572	17.2	739	10 O9LEF8	09lef8 arabidopsis
41	571.5	17.2	609	10 O9C8W6	09c8w6 arabidopsis
42	571.5	17.2	662	10 O949Y4	0949y4 arabidopsis
43	571	17.2	609	5 O9VQNA	09vqna dirosophila
44	567	17.0	626	5 O8T684	08t684 dictyostell
45	565	17.0	1528	5 O8T677	08t677 dictyostell

ALIGNMENTS

RESULT 1

ID OBR543 PRELIMINARY; PRT; 673 AA.

AC OBR543; 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Sterol in 2.

GN ABC8.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=129/SV;

RA Lu K., Zhou Y., Lee M.-H., Patel S.B.;

RT *Molecular cloning, genomic structure and characterization of novel

mouse head-to-head tandem ABC transporters.*;

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF351811; AAL82898.1; JOINED.

DR EMBL: AF351799; AAL82898.1; JOINED.

DR EMBL: AF351800; AAL82898.1; JOINED.

DR EMBL: AF351801; AAL82898.1; JOINED.

DR EMBL: AF351802; AAL82898.1; JOINED.

DR EMBL: AF351803; AAL82898.1; JOINED.

DR EMBL: AF351804; AAL82898.1; JOINED.

DR EMBL: AF351805; AAL82898.1; JOINED.

DR EMBL: AF351807; AAL82898.1; JOINED.

DR EMBL: AF351808; AAL82898.1; JOINED.

DR EMBL: AF351809; AAL82898.1; JOINED.

DR EMBL: AF351810; AAL82898.1; JOINED.

SO SEQUENCE 673 AA; 76008 MW; FA08340445DF259C CRC64;

Query Match 21.2%; Score 704; DB 11; Length 673;
Best Local Similarity 28.7%; Pred. No. 5.8e-43;
Matches 195; Conservative 130; Mismatches 261; Indels 94; Gaps 17;
11 GSKGLQVNRGSSSLGAPATAPDEPHSGILHASYSVSHRVR-----PMWD-I 57


```

Db      14 GTVLADASGGLDSSL-----FSESDNSLFTYSGGSMLEVDLTVYONDIAQVWFEDL 69
Oy      58 TSCROOFTROI-----LKDVSLYESGQIMCIGSSGSKTLLDMSGRILRACGF 109
Db      70 AQFKIPRSHSSQDCELGIRNLKFRVSGOMALITGSSGCCRASLDVYITER-CHGGM 128
Oy      110 -LGEYVNGRALRREQFQDCSYVLQSDPTLLSLTVRETIHYTALLAI-RGNPGSPORK 167
Db      129 KSGQWINGOPSTPOLVRRKCAVHVRHQDOLLPLNLTRETIYAQMRLPRFSQAOROKR 188
Oy      168 VEAVAAELSLSHVADRILGNYSIGISTGERRRVSIQAOLLODPVLMFDEPTGLDCMT 227
Db      189 VEDVIAELRLQCCATRGVGMNTYVRGSGERRRVSIGVOLMNPICILLIDEPTSGLDSET 248
Oy      228 ANQIVLVLELARRRIVLVLTIHOPRSELPOLFDITATISFEGELJFCGPAPMLPFENDC 287
Db      249 AHNLTITSLRLAKGRNLVLISLHOPRSDIFRLFDLVLTMTSGPTITYLGAQOQVOYFISI 308
Oy      288 GYPCPEHSPDFDYMDLTSVDTSKEREIETSKRYOMESAYKSA-----ICHKTLKN 341
Db      309 GHPCGRYNPADEYVDLTSDIRSKEREVAATVEKAQSLALELEKVGDFDPLMKAEKE 368
Oy      342 IERMHLKTLPMVPRKTD-----PGVPSKIGVLLRRTNLRNKLAVITRLION 393
Db      369 LMTSTHVSLETL-----TQDTCGTAVELPGMLEQESTLIRQISNDPFDLPTLLHSGSA 424
Oy      394 LIMGILFLFVLRYVNSVLKGAIDORVGLTQFVATPYTGMLNANVLEPVLRAVSDS 453
Db      425 CIMSLLIGFLYXGHAKOL--SFMDTALLFRMIGLIFPNVILDVVSKCHSERSMYEL 482
Oy      454 QDGLYOKQMMLAVLHVPFSVATMTFSSVCYWTGLIHEVARNFYSAALLAPHLIG 513
Db      483 EDGYTAGPYFAKILGELPEHCAVILYAMPYVLTMLRP-----VPELFL 529
Oy      514 EFLTVILGIYONPRIVNSVALS-----IAGVLYGS-----GFLRNIOEMPPIPKI 561
Db      530 LHTLVMLVPCRCRMALASAMLPFMSSFFCNALNSFYLAGFMINDMLNMIYAVM 589
Oy      562 ISYTFORKESEILVNVNEFYGLNFT--CGSSNVSTVTPMCAFTQGIQIEKTCGATSR 619
Db      590 ISKLSFLMKCSGLMQIOFNGHLYTTOIGNFTPSILGDTM-----ISA 632
Oy      620 FTMNPLIYSFIPALVIIGI 639
Db      633 MDLNSHPLYAYT--LIVIGI 650

RESULT 2
Oy      096TAB PRELIMINARY; PRT; 655 AA.
AC      096TAB;
DT      01-DEC-2001 (TREMBLrel. 19, Created)
DT      01-DEC-2001 (TREMBLrel. 19, last sequence update)
DE      01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE      ATP-binding cassette superfamily 6 (White) member 2.
GN      ABCG2.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
MEUline=21201983; PubMed=11306452;
RA      Komatsu H., Kotani H., Hara Y., Nakagawa R., Matsumoto M.,
RA      Arakawa H., Nishimura S.;
RT      "Identification of breast cancer resistant protein/mitoxantone
RT      resistance/placenta-specific, ATP-binding cassette transporter as a
RT      transporter of NB-506 and J-107088, topoisomerase I inhibitors with an
RL      Cancer Res. 61:2827-2832(2001).
DR      EMBL; AB051855; BAB46933.1;
DR      InterPro; IPR003439; ABC_transportr.

```

```

DR      InterPro; IPR003880; Pantene_atlch.
DR      InterPro; IPR003580; Prochaykinln.
DR      Pfam; PF000005; ABC_tran.1
DR      ProDom; PDOM00006; ABC_transportr; 1.
DR      SMART; SM00203; TK; 1.
DR      PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_1.
DR      ATP-binding.
SQ      SEQUENCE 655 AA; 72314 MW; ABAF66B96034C5A8 CRC64;

Query Match      20.5%; Score 680.5; DB 4; Length 655;
Best Local Similarity 29.2%; Pred. No. 2.9e-41;
Matches 182; Conservative 137; Mismatches 250; Indels 55; Gaps 18;

Oy      21 SSSLEGAPATP---EPHSIGLIHASYSVHVRPMMWDINSCROOFTROIADKSLAYE 77
Db      13 SGGNNGPPATASNDLKFTBGAVALSFNNICRVALKSGFLPCRPRVERKILNNGIMK 72
Oy      78 SGQIMCIGSSGSGKRTLLDMSGRILRACGFLEGVYNGRALRREQFQDCFSYVLQSDT 137
Db      73 PG-LNALIGPGGGKSSLLDVLAAKDPG--LSGDVLLNG-APRANKKNSGYVVDV 129
Oy      138 LLSLTVEETLHYTALLAIRGNPG-SFOKVEAVMAELSLSHVADRILGNYSIGISTG 196
Db      130 VMGTLVRENLOFSALRLATMTMTNHEKNERINRYIOELGLDKVADSKVGTQFIRGSG 189
Oy      197 ERRRVSIQAOLLODPKVLFPDEPTGLDCMTANOIVLVLELARRNRIYVLTIHOPREL 256
Db      190 ERKRSTIMEITDPSILFLDEPTGLDSTANAVLLILKRRSKOGRTIISHIQPRSI 249
Oy      257 FQLEDKIALISFEGELIFCGTPAEMIDPFNDGYPCEPHSNPDFYMDLTSVDQO----SK 312
Db      250 FKLFDLSLLASGRLMFHPGPOALGYESAGHCEAANNADFDLIINDSTRVALNR 309
Oy      313 ERE-----IETSKR---VOMIESANKSAICHT-----LNIERMKHLKTLPMVPE 356
Db      310 EEDERATEIIEPSSODKPLIEKLAIEYVNSFYEKRAELHQLSGEKKKTIYFEKEISY 369
Oy      357 KTKSPGVSFKIGVLRVTRNLRNKLAVITRLIONIMGLF--LEFVRYVNSVLK 414
Db      370 TT-----SCHOLRWVSKRSFNLGNPOASTIAQITVYVLDLVGATFGLKNDST---- 421
Oy      415 AIDORVGLLXQVATPYTGMLNANVLEPVLRAVSDSODGLYOKQMMLAVL-HVLP 473
Db      422 GIORAGVLYF-FLFTNOCFFSVSAVELFVEYKELFHEIYISGYRVSSYFGLKSLDLP 480
Oy      474 FSVVATMTFSSVCYWTGLIHEVARNFYSAALLAPHLIGELTVLGIYONPRIVNSV 533
Db      481 RMPLPSIITFCIVYFMGLKPKADAFVMMFTLM--MVAYSASSMALAIAAGOSVSA 537
Oy      534 VALLSIAGV--LVSGFLRNIOEMPPIPKIISYTFORKESEILVNVNEFYGLNFTGSSN 591
Db      538 TLMFTICVFMIRSGILLVNTTITASMLSLQYFSIRICFTALQHNHEFLQNFPCG--- 594
Oy      592 VSVTTPMCAFTQGIQIEKTCG 615
Db      595 LNTAGNPNCAV-----TCTG 610

RESULT 3
Oy      096LD6 PRELIMINARY; PRT; 655 AA.
AC      096LD6;
DT      01-DEC-2001 (TREMBLrel. 19, Created)
DT      01-DEC-2001 (TREMBLrel. 19, last sequence update)
DE      01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE      ABC transporter ABCG2.
GN      ABCG2.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.

```

RA Schuetz J.D., Wall A.M., Sampath J., Sorrentino B., Du G.;
 RT "The Human ABC Transporter, ABCG2, Transports Hoechst 33342 and
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY017168; AAG52982.1;
 DR InterPro: IPR003439; ABC_transporter.
 DR InterPro: IPR003880; Pantoate_attach.
 DR Pfam: PF00005; ABC_tran; 1.
 DR Prodom: PD000006; ABC_transporter; 1.
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN.1.
 SO SEQUENCE 655 AA; 72288 MW; B3B5D0C02C095CA48 CMC64;

Query Match 20.28; Score 672.5; DB 4; Length 655;
 Best Local Similarity 29.08; Pred. No. 1.1e-40;
 Matches 181; Conservative 137; Mismatches 251; Indels 55; Gaps 18;

QY 21 SSSLEGAPATAP---EPHSLGILHASYSVSHRYRPMMDITSCROMTROILKDVSLAYE 77
 Db 13 SSGTNGFPATASMDLKAFTGCAVLSFHNICYRKLKSGELPCRKPEVEKILSNINGIMK 72
 QY 78 SGOIMCITIGSSGSKTLLDAMSGRLRAGTFLGEEVYNGRALTREOFODCFESYVLQSDT 137
 Db 73 PG-LNAILGPTGGKSSLLDYLAARKDPGSG-LSGDVLING-APPRAFCKSGSYVDDV 129
 QY 138 LLSLYRETLHYTALLAIRGNPG-SFOKKVEAAMELSHVADELIGNYSIGISTG 196
 Db 130 VMGLTYRENLQFSALRLATYTMNHEKNERINRIQELGLKVAADSKVGQIFRGVSG 189
 QY 197 ERRRVSAIADLDOPKMLDFDEPTGDCMTANOIYLVLEARNIVLTIHQPSSEL 256
 Db 190 EKRRTSGMELTIDPSTLFLDEPTGDSSTANAVILLLRMSKOGTITFISHQPRYSI 249
 QY 257 FOLEDKAITSFGLIFCGTAPAEMLDFENDCGYPCPEHSNPDEYMDLTSDVQ----SK 312
 Db 250 FKLFDSTLLASGLMHGPAQELGYFESAGYCEAYNNPADFLLIDGSTAVALLNR 309
 QY 313 ERE-----IETSKR---VOMIESAYKSAICHT-----LKNIRMKHLKTLPMVPF 356
 Db 310 EEDFKATEILEPSKODPLLEKLAETIVNSPFKETAELHQLSGCKKKKITFYKISY 369
 QY 357 KTDSQGVSKLGLARVTRNLYRKNLAVITLQNLIMGLFL--LEFVLVRASNYLKG 414
 Db 370 TT-----SECHQLRWVSKRSRKNLGNPOASIAOITIVYGLVGAIFYGLKNDST---- 421
 QY 415 AIDRVGLLYQFAGATPYTGMLNAVNLFPVLRAVSDQSDGLYOKNMMLAYAL-HVLP 473
 Db 422 GIONRAGVLF-ELTNOCSESSVAVLEVEKKLEFHEIYSGYRVSSTYFLKLLSDLP 480
 QY 474 FSVVAWIFSSVCYWTGLHPEVARFGYFSALLAPHLIGEFLLVLGLIQVNPNTVNSV 533
 Db 481 MRMLPSTIFTCIYFMYHGLKAKADAFVMMFTLM---MVAYSASSMALALAAQSYVSA 537
 QY 534 VALISTAGV--LVSGGLRNIOEMPIPKIISTFYTKYCSLIVNEFYGLNFTCGSSN 591
 Db 538 TLMATICFVPMIIFSGILVNLITIASNLSWLOFISIPRYGTALQHNFIQGNRCPC--- 594
 QY 592 VSVTNPMAFTGIGIETKCPG 615
 Db 595 LKATGNPCMYA-----TCTG 610

RESULT 4

Q8T691 PRELIMINARY; PRT; 801 AA.
 AC Q8T691;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DE ABC transporter Abcgl. 21, last annotation update)
 GN ABCgl.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
 NCBI_TaxID=44689;

RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-AX4;
 RA Aujard C., Loomis W.F.;
 RT "Evolution of the ABC Transporters of Dictyostelium";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF482380; AA191485.1;
 SO SEQUENCE 801 AA; 90052 MW; CCC4F003CB195A3 CMC64;

Query Match 19.98; Score 662; DB 5; Length 801;
 Best Local Similarity 27.48; Pred. No. 8.3e-40;
 Matches 185; Conservative 134; Mismatches 246; Indels 110; Gaps 16;

QY 61 ROQWTRQILKDVSLAYESGOIMCITIGSSGSKTLLDAMSGRLRAGTFLGEEVYNGRAL 120
 Db 131 KKKISKQILTNINHIESGTFALMGPSGACKTLLDLARLHINIS--GTMYLNGKKS 188
 QY 121 RREOFODCFSYVLQSDTLLSLAYRETLHYTALLAIRGNPGSFQ-KKEVAMELSH 179
 Db 189 DENIFKICGYVTQSDSLMPSLTVRETLNFAQLKMPDVPFLKELARVODIIDEMGLNR 248
 QY 180 VADLLIG--NLSLGISTGERRRYSIAAQLQDPKMLFDEPTGDCMTANOIYVLVE 237
 Db 249 CADTLVGTADNRKIGISGGERRRYTISSIELTGFSVILLDEPTGLDASTSFYVMSALK 308
 QY 238 LARNRRLVLTIHOPRSELQLPFKIATLSFGLIFCGTAPAEMLDFENDCGYPCPEHSNP 297
 Db 309 LAKSGRTITICTIHOPRSNITDMPDNLILDGNTIYKANKALEYFNANGYHCEKTNP 368
 QY 298 FDFYMD--TSVDTQ-----
 Db 369 ADFELDLINQVEQADSDDDVDNDEEEIGGGGGGCGAGTIEDIGISPTMNSAV 428
 QY 312 ---KERIE-----
 Db 429 DNIRNNELKQD 488
 QY 342 IER-----MKHLKLPVPPFKTNSPQVFSLSGLVLRVNRNLYRKNLAVITLQNLIM 396
 Db 489 ISKENRTDFEYK-----RGPNELTQFSLDGLREVYNARHMAFVNLQAIQIFQ 539
 QY 397 GLFELFVLRASVWLKGAIDRVGLLYQFAGATPYTGMLNAVNLFPVLRAVSDQSDG 456
 Db 540 G--LLCGIYVYQGLGSSVSQSKGVAAFIIMGSPFAVNSTHVPDVTITFLKDRSG 597
 QY 457 LYOKNMMLAYALHVPFSVVAWIFSSVCYWTGL--LHPEVARFGYFSALLAPHLIGE 514
 Db 598 VDTLPPEFLAKSPFDACIAVLPLPVATFIYMMNQNQVDPYSAPAFRRFLM---LVLA 654
 QY 515 FULVILGIYQN---PNI--VNSVALLSTAGVLSGFLRNIOEMPIPKIISTFYTKY 570
 Db 655 SQTCLSIGLVLISSVPMVQGTAVAPLIVILFELFSGFILNLDVPMGLWFPYISFERY 714
 QY 571 CSELIYVNEFYGLNFTCGSSNSVSTNPMCAFQGTGIGIETKCGARSPTMNLILYSE 630
 Db 715 MIEAAVYNAKDVHFTCTDQ---KIGGCPVOYGNVNI--NGGYDIDHFRNVMILVLY 770
 QY 631 IPALVILGIYVFKIR 645
 Db 771 IIGFRVLTFLVCLKK 785

RESULT 5

Q9R004 PRELIMINARY; PRT; 657 AA.
 AC Q9R004;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DE Breast cancer resistance protein 1.
 GN ABCG2 OR BCRP1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid-10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-FVB; TISSUE-LIVER;
 RX MEDLINE-99413474; PubMed-10485464;
 RA Allen J.D., Brinkhuis R.F., Wijnholds J., Schinkel A.H.:
 RT "The mouse Bcrp1/Abcp gene: amplification and overexpression in
 cell lines selected for resistance to topotecan, mitoxantrone, or
 doxorubicin."
 RL Cancer Res. 59:4237-4241(1999).
 DR EMBL: AF140218; AAD54216.1; -
 DR MGI:1347061; Abcg2.
 DR InterPro: IPR003439; ABC_transport.
 DR InterPro: IPR003880; Ppantne_attach.
 DR Pfam: PF00005; ABC_tran; 1.
 DR Prodom: PD000006; ABC_transport; 1.
 DR PROSITE: PS00012; PHOSPHOPANTHEINE; UNKNOWN.1.
 SQ SEQUENCE 657 AA; 73021 MW; 207870BC272CCD5 CRC64;

Query Match 19.8%; Score 660; DB 11; Length 657;
 Best Local Similarity 28.0%; Pred. No. 9e-40;
 Matches 181; Conservative 135; Mismatches 242; Indels 88; Gaps 19;

OY 13 MGIQVNRGSSQSLGAPAPAPPHSGILHASYSHRRPMDITSCROQTROLTKDY 72
 DB 12 MCRNRNNGLPNRNSRVRFLAEDVLSFHITTYR--KVSGLV--RTYKEKELISDI 66
 OY 73 SLVESGQIMCIIVSSSGSKTLLDMSGRLGAGTFLGEVYNGALRREODPESYV 132
 DB 67 NCIMKRG-LNALITGPTGGKSSLDVLAARKDPRG-LSGDVING-APPAHFCSCGV 123
 OY 133 LOSDILLSLVREILHTALLAIRGNPSFO-----KVEVAEMLSLSHVADRLGN 187
 DB 124 VDDVVMGTLVRENIQFSAALRL--PTMKNEKNERINTIIELEGEVADSKVGT 179
 OY 188 YSLGISTGERRRVIAQLDOPKMLDEPTGTGCMFANOIVLYELARRNIVLV 247
 DB 180 QTRIGISGGERKRTSIGMELITDPSLFDPTGTGDSSTANAIVLLLRMSNGRTIIF 239
 OY 248 THQPRSELQDLFDKIAILSPGELIFCGTPAEMLDFNDCGYPCEHSPNPFYMDLTSV 307
 DB 240 SIHQRYSTFKFLDSLITLASKLVHGAOKALEFASAGHCEPYNNPADPFDVING 299
 OY 308 DPGS-----KEREIEFSR-----VOMIESAYKKAICHTKLNIEKKHLKTL 352
 DB 300 DSSAVMLNREEDONEANKTEEPSKGEKPYENLSEFYINSAIYGETKALDOL----- 352
 OY 353 HVPFKTDSGVFSKLV-----LARRVRNIVRNKLAVITRLONL 394
 DB 353 -----PGAOEKKGTSAFKEPYVYSFCHQLRMIAKRSKLNLLGNFOASVAOLITV 403
 OY 395 IAGLEFL--LEFVLAVRSNVLKGAIDRVGLLYOVGAPPTGMLNANVLEPILRAVSDE 452
 DB 404 ILGLIIGAIYFDLTKDA---AGMNRAGVLF-FLITNCCFSYSVAVELFYEEKLFIHE 458
 OY 453 SDDGLYOKRMALAL-HVLFPSVAATMIESVCTWTGLHPEVARFGYSALLAPHL 511
 DB 459 YISGYRVSSTYFEGVMSDLPMPRLPSYIFCIYFMGLKTKYDAEFIMAFITL--M 515
 OY 512 IGEFTLVLTGIVONPNIYNSVALLSTAGV--LVGSGFLRNIOEMPIFKIISTFOK 569
 DB 516 VAYTSSMALAIATQSVAATLTATIAFVFMFLSGSLVLRIGPMLSWLOYSIPR 575
 OY 570 YCSEILVYNEFGINFTCGSSNVSTINPMCAFTGOGIOFIERTCG 615
 DB 576 YGFTALOYNEFLGGEFCPG---FNTDNTGCVNSYAI-----CTG 612

RESULT 6
 ID 09M3D6 PRELIMINARY; PRT; 725 AA.
 AC 09M3D6;

DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE ABC transporter-like protein (putative ABC transporter protein).
 GN T26112.10 OR A3G55130.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_Taxid-3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Monfort A., Casacuberta E., Puigdomenech P., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X., Queller F., Salanoubat M.;
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Goldsmith A.D.,
 RA Jiang P.X., Lee J.M., Onodera C.S., Quach H.U., Tang C., Toriumi M.,
 RA Yamamura Y., Yu G., Yu S., Bowser L., Carlinici P., Chen H., Cheuk R.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamuya A., Karlin-Neumann G.,
 RA Kawai J., Kim C., Kosemura E., Lam B., Lin J., Meyers M.C., Miranda M.,
 RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
 RA Shinn P., Southwick A., Tracy S.E., Shinozaki K., Davis R.M.,
 RA Ecker J.R., Theologis A.;
 RT "Full length cDNA of gene T26112.10/A3G55130 (GI:7019646)."
 RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Eguu P., Lee J.M.,
 RA Toriumi M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,
 RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,
 RA Southwick A., Davis R.W., Ecker J.R., Theologis A.;
 RT "Arabidopsis Open Reading Frame (ORF) Clones."
 RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AL132954; CAB75747.1; -
 DR EMBL: AY045932; AAK76606.1; -
 DR EMBL: AY079387; AAL85118.1; -
 DR InterPro: IPR003593; AAA_Atpase.
 DR InterPro: IPR003439; ABC_transport.
 DR Pfam: PF00005; ABC_tran; 1.
 DR Prodom: PD000006; ABC_transport; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN.1.
 DR ATP-binding.
 KW KM
 SQ SEQUENCE 725 AA; 80656 MW; 790C535A7929CC16 CRC64;

Query Match 19.5%; Score 649.5; DB 10; Length 725;
 Best Local Similarity 29.4%; Pred. No. 6e-39;
 Matches 182; Conservative 124; Mismatches 246; Indels 67; Gaps 15;

OY 33 PEPHSIGILHASYSHRRPMDITSCROQTROLKDVSLVESGQIMCIIVSSGSK 92
 DB 68 PVYVYLNFRNNLDYDTLRR-----FGFSQNRNVTKLDDVSEADCDITANVLGSGACK 123
 OY 93 TLLDMSGRLGAGTFLGEVYNG-RALRREQDFDCFSYVLAOSDTLLSLVRETLHT 151
 DB 124 STLDALAGRYAE-GLSGSYVINGEKVLAQSLTKVISAAYVQDDLLPMLTKVETLWFA 182
 OY 152 ALLAIRRG-NGSPQKRYEAVMAELSLSHVADRLGNLSLGISTGERRRVIAQLD 210
 DB 183 SEFRLPRLSKSKMERREALIDGLNRNANTYVIGDGHGVSQGERRRVSIIGDIIHD 242
 OY 211 PKVMLDEPTGTGCMFANOIVLVLELARRNRYVLTIRHPSLFPOLFKIILSPGE 270
 DB 243 PIVLEDEPTGISTNFMVVOYLAKRIAGSGSYIWSIHPSARIYVELDRLIILSNCK 302
 OY 271 LIFGCTPAEMLDFNDCGYPCEHSPNPFYMDLTSVDTQSKEREIETS----- 319

```

Db 303 SVENGSPASLPGEFSDGRIPEKENISERALLV-----RELEGSNEGTAALVDN 354
QY 320 -----KRVOMIESAYK-----KSAICHKVL--KNIERMKHLKTLPMVPEKTKD 360
Db 355 ERMQONKISLIQSAPOKNDLDDRSLSLKEAINASVSGKLVSSSSNPTSEVSYA 414
QY 361 SPVEFSKLVLLRRVTRNLVKNLAVITRLONLIMGLFLFVLVRAVSNVLCATODRV 420
Db 415 NPSLFEFE-ILARRKMKMTRMPELVOTRIATVAVTGC-LLATVYMKLDHTPRGA-QERL 471
QY 421 GLLYOFGATPYTGMLNANVLFPVLRAVSDQESODGIXOKMMLAALHVPESVATM 480
Db 472 -TLFATVYPTMFCCLDNVPEVFOERTFLRETHNAVRYSSVYISHSLVLPQLAPSL 530
QY 481 IFSSVCTWTLGLHPEVAREGFSAALAPHLIGEFLTLVLGIYQNPNIYNS--VVALLSI 539
Db 531 VFSATIFMTVGLSGGLEGFVYCLLIYASFWSSSVTFISGV--PMLLCYVASTITYL 588
QY 540 AGVLVSGFLRNIOEMPIPEKTIISYTFPOKCEILVNEPFGMLFTCGSSNVSVTTPM 599
Db 589 ACCLISGFEVNDRIPEFTWTFMTHYISILKPYEAVLINEF-----DDPS 633
QY 600 CAPTGOIOFEKPCGATS 618
Db 634 RCTVRGVQVFDSTLLGVS 652

RESULT 7
Q8RM19 PRELIMINARY: PRT: 691 AA.
AC Q8RM19;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 77.2 kDa protein.
GN AT3G21090.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosida II; Brassicales; Brassicaceae; Arabidopsids.
ON NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carriol P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinzaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RA Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY093054; AAM13053.1;
DR Hypothetical protein.
SQ SEQUENCE 691 AA: 77219 MW: 6E473CC0B440D7E9 CRC64;

Query Match 18.9%; Score 628; DB 10; Length 691;
Best Local Similarity 28.1%; Pred. No. 2,1e-37;
Matches 173; Conservative 123; Mismatches 225; Indels 94; Gaps 17;

QY 25 LEGAATAPAE-PHSGILHASYSVSHRVPMDITSCROOQ---TRQILKQVSLYVESG 79
Db 3 LEGSSSGRRQLPSKLEMSRGAYLA-----WEDLVIVVPFSDGPTRLRLQNLNGAEPG 56
QY 80 QIMCLIGSSGSKTILLDAMSGRLGRAGTFLGEEVYVNGRALREDFQDFSVVLOSITL 139
Db 57 RIMALMGPSGSKTILLDAMSGRLGRAGTFLGEEVYVNGRALREDFQDFSVVLOSITL 114
QY 140 SSVLRRETLHYTALLAIRGNGSPQOK-----VEAVNAELSLSHVADRLIGNSLGIGS 194
Db 115 GLTVARETITYSAHLRL-----PSDMKSKEVSDIVSGTITIELGLDQCSDRVIGNMARGVS 170
QY 195 TGERRRVSIAMLODDPYMFLDEPTGIDCFTANQIYVLLVELARRNIVVLTTHOPRS 254
Db 171 GGERRVSAILEILTRPOLFLDEPTSGIDSASAFVIOALNINIRADRTVISVHOPSS 230

```

```

QY 255 ELFOLEPKIALISFEGLEFCPTAEMDFENDGCGPCPEHNSPDPFMDLTSDVDSOKR 314
Db 231 EVAFLEDDLEFLSSGESVTFGEAKSAVEFEAGSGFPCKPRKNSDHLRCINDPDTVA 290
QY 315 EIEFSKRVQ-----MIESAYKSAICHKTLKNIEMKHLKTLPMV 354
Db 291 TLKGSQRIOETPATSDPIMNLATSVIKARLVEN-YRKSRAKSAKSTRLSNIEGLEME 349
QY 355 PEKTSQGVFSKLVLLRRVTRNLVKNLAVITRLONLIMGLFLFVLVRAVSNVLCG 414
Db 350 IRKGSSEATW-WKOLRLTLARSEFYIMCSDVGYWTRITSIYVSI-----392
QY 415 AIDRGLLYOFGATPYTGMLNANVL-----FPVL---RAVSDQESODG 456
Db 393 ----SVGTIFYDGYG-YYSILARVSCGGFTTGMFMMSIGGFPSPSEEMKVFYKERTSG 447
QY 457 LYOKMOMLALALHVPESVATMTMFSVVCWTGLGLHPEVAREGFSAALAPHLIGEFL 516
Db 448 YGVSVYIILSNYISSPFLVAISVITGTYINLYKRRPGSHAFCLNIFSVSYESL 507
QY 517 TLVLGIYQNPNIYNSVALLSING-VLVSGFLRNIOEMPIPEKTI-----ISYTFQKY 570
Db 508 MMYVASV--PNFLMGLITGAGLIGIMMISGFERLLPDLF--KIFMRYPSYISGWS 562
QY 571 CSRLVNVNEFYGNF 585
Db 563 AIOGGYKNDPLGLEF 577

RESULT 8
Q9ARU4 PRELIMINARY: PRT: 668 AA.
AC Q9ARU4;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative ABC transporter.
GN P0445D12.3.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoae; Oryza.
ON NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-CV. NIPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RA "Oryza sativa niponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0445D12.3."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AP003046; BAB40032.1;
DR InterPro: IPR003593; AAA_Arase.
DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
DR ATP-binding; Transport.
SQ SEQUENCE 668 AA: 73368 MW: D1875B8C75B0F3B2 CRC64;

Query Match 18.7%; Score 622; DB 10; Length 668;
Best Local Similarity 31.0%; Pred. No. 5,4e-37;
Matches 171; Conservative 100; Mismatches 245; Indels 36; Gaps 12;

QY 55 WDITSQ-----RQMTROLKQVSLYVESGQIMCLIGSSGSKTILLDAMSGRLGRAGT 108
Db 58 WARTICALKNRKGDVARELLSNASGEAKSGRLALMGPSGSKTILLANVLAQOLVASPSL 117
QY 109 -FLGEEVYVNGRALREDFQDFSVVLOSITLSSLTRETLHYTALLAIRRG-NGSPQK 166
Db 118 HLSGFLYINGRIRISGGRK--IAYVRQEDLFFSOLVRETLSLAAELQRLRTPERRRS 175

```

```

RESULT 9
09L182
ID 09L182 PRELIMINARY; PRT; 672 AA.
AC 09L182;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ABC transporter-like protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Brassicaceae; Brassicales;
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kameo T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Nakamura Y.;
RT 'Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ntney pl,
RT RAC and BAC clones.';
RL DNA Res. 7:217-221(2000).
DR EMBL; AP001313; BAB03081.1; -.
DR InterPro; IPR003593; AAA_Atpase.
DR InterPro; IPR003459; ABC_transport.
DR InterPro; IPR003860; Pntane_attach.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transportr; 1..
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_1.
ATP-binding.
SQ SEQUENCE 672 AA; 75269 MW; 20B2D99215600135 CRC64;
KM

```

Query Match: 18.6%; Score 618.5; DB 10; Length 672;
Best Local Similarity 32.1%; Pred. No. 9.8e-37;

Matches 182; Conservative 100; Indels 79; Gaps 17,

OY RQILDVSLIYSSGGIMKILSSSGGRTTLLDMSGRGLRGATFLGEAYNGR---ALR 121
| : | | | : : : : | | | | | : | | |
DB RLVLKCVSGIYVPGLMLMGPSGGKTTLVTALAGRL--OGKLSCGVSYNCEPTTSVK 154

OY 122 REOFODCRSVYLQSOTLSSLTVRELTLYHTLLLAIIRGNPSPFOK-----VEANMELS 176
Db 155 KRT-----GFYTQDVLPPLHLMETTYTTALLRK----PEELRNKELEOVEMAVSDIG 205

OY 177 LSHVDRIIGNSTLSIGISTGERRRYSIAOLLODPKVAFDEPTTGIDCMATANQIVLV 236
Db 206 LTROCSNYVGGLLRGISGERKKRVSIGEMLVNPSILLIDEPTSGLDSTAARIVATLR 265

OY 237 ELARNKRVIYLIHQPRSELVOLPKKIILSFGLIFCGTAEMLDFFNDCCGY-PCPHHS 295
Db 266 SLARGGRVYTHIQPRSRLYMFPKYVLTSEGCPIYGDSGRVMFYEGSIYGGSFVF 325

OY 296 NPFDYMDLTSDPDOSKER-EIEESKRVOMIE-----SAYKSACIKTKNIKIE 343
Db 326 NPADEVLDLANGITSDITYQTQDIETNGHLDRLEONISVKGSLISSYKN--LYPLEKE-- 381

OY 344 RKMKLKTLPMWPFKKDSPGV-----FSKLGVLLNRVRNLVNRKLAVIT 388
Db 382 --EVSRTPP-----QQDTNARKRKATTNNRPMTSMWGFSVLK---RGLKERSHESES 430

OY 389 RLQLNWINGLFLLFEVNVARSNTAKGAIODRVGLLYOFVGTAPPTYGMNAVNPFLPVRAV 448
Db 431 GLRFPMVASVSILSGLLMMHRSRV-AHRDDOVGLIFFEPSIPFWGPFLPNALFTFPOERPM 488

OY 449 SDOESGDGLXOKQMOMLAVALHYDFSVVAAMITSVCYWMLGLHPFAARGYSAAILLA 508
Db 489 LIKESSGSIIYRSSTYYARTFGDDLPEMIILTFITVTYYYMGSKPSTTEITWTMLTVLY 548

OY 509 PHLGEEFTLVLLGIVOPDNINVSVALLSIAGVYSGSGFRNIQEMPPIPKITSYSTFO 568
Db 549 NVIVAQGAGIALGAILDMARKAALTSSVLMVPLLAGEY---IQHIPGFIAMLKYSIFS 605

OY 569 KYCSEILVANEPYGLNF-----CGS 589
Db 606 HCYCKILV-----GWQYTWDEUYECGS 627

RESULT 10

OGENBS PRELIMINARY; PRT; 727 AA.

ID OGENB5 AC QGENB5; ID OGENB5; AC QGENB5;
DB 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)

Df ABC transporter-like protein.
De Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsils.
OX NCBI_TaxId=3702;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=COLUMBIA;
RX MEDLINE=98069011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N., Tabata S.;
Sequence analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen physically assigned pl clones." ;
RL DNA Res. 4:291-300(1997).
DR EMBL: AB006704; BAB08684.1;-
DR InterPro: IPR003593; AAA_ATPass.
DR InterPro: IPR003439; ABC_tranpass.
OR Pfam: PF00005; ABC_tran; 1.
OR ProDom: PD000006; ABC_transportr; 1.

Db 508 MARYASV--PNTLMLTGTGAGLIGITMTSGFFRLDLPKIFWRYPVSYISYQMAIQ 565

QY 571 CSEILVNEEYGLNETC 587

Db 566 AS-LRSGSKAGESRTEC 581

RESULT 14

QY 09M2V7 PRELIMINARY; PRU; 720 AA.

AC 09M2V7

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE ABC transporter-like protein.

GN T1SC9.80.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.

OC NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Barques M., Collado M.C., Navarro P., Terol J., Perez-Alonso M.,

RA Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X., Quettler F.,

RA Salanoubat M.,

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Arabidopsis sequencing project;

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AL132970; CAB82704.1;

DR InterPro: IPR003593; AAA_ATPase.

DR InterPro: IPR003439; ABC_transporter.

DR Pfam: PF00005; ABC_tran; 1.

DR ProDom: PD000006; ABC_transporter; 1.

DR SMART: SM00382; AAA; 1.

DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOMN_1.

DR ATP-binding.

SO KW

Query Match 720 AA; 79889 MW; 9D15431098CF1A3 CRC64;

Best Local Similarity 18.5%; Score 614; DB 10; Length 720;

Matches 182; Conservative 122; Mismatches 254; Indels 76; Gaps 17;

QY 23 SSLEG--APRTAPEPHSIGLTHSYVS-----HVRVPMWDTSCROOQTQOILKDV 73

Db 40 SSIDGNDHLMRPVPLSFNNLTYNVSRKILDFHDVLMVRRTSFSK--TTLLDNIS 96

QY 74 LYESGOIMCIISSGSGKTTLLDMSGRLAGTFLGEYVVGRLARREOFODCSYVL 133

Db 97 GTRDEDELLAVLGAASSGSKSTLDALANRAK-GSLKGYTLANGELQSMKLVISAYVA 155

QY 134 QSTLLSSLTAVRETLHTTALLAIRKNPSSQF-KYEAVALSELSHVADRLIGNYSLG 192

Db 156 ODDLFPLMVEETLFAAEFLRPLRSLPKSKKLROVALIDQIGIRNAATIIIGDSHG 215

QY 193 ISNGERRRYSIAQQLDPRVLMFDEPTGLDCTANQIVLVLELARRRYVLTIHOP 252

Db 216 ISGERRRYSIGIDIIHDPVLFLEDEPTGLDSTSAFMYVKVLKRIAESGSIITIMSIHOP 275

QY 253 RSELFOLFRIALISFGEILFCSTPAEMLDFFNDGCPCEHSPNPFDEVADLNS----- 306

Db 276 SHRYVSLDLRLITLISGHYVFSASPISLSEFAGFPNPPIENNOPEFALDLRELGS 335

QY 307 -----VDTOSKREIETSKRVOMIESAVKSAICKTLK-----NIERKHLK----- 349

Db 336 GGTGRIYEFNKWKQEKOSNPQTLTPPASPNP--NLTLKEAISISISGKLVSGGGGS 393

QY 350 -----TLMPWPFETKDSPGVFSKGLVLRVYTNLVNKLAVITRLQNLIMGLFEL 401

Db 394 SVINHGGLTAVAFANP-----FWIEIKTLTRRSILNSRQPELLQMLATVIVTG-FILI 448

QY 402 FFVLKRVSNVLKGAIDORVGLLYOFVATPYTGMILNVLFPVLRAVSDQESODGLYOKW 461

Db 449 ATFWWRIDNSPKG-VQERLG-FEFAFASMTMYTCADALPVLQRIYFMRETAANARRS 506

QY 462 QMLATVLAHLPEFSVATMIFSSVCYVTLGHAPVARRGYFSAALLAPHLIGELTVLV 521

Db 507 SYVLSHAIVTFPSLIFLSLAFVATTFVAVGLEGLMGFLFCLIIILASFMSGSSEFTFLS 566

QY 522 GIVONPNIV--NSVALLSLAGVLSGFLRNQOEHPKILSYTFPOKCEILVNV 578

Db 567 GVV--PVMGLYITVAIL--AFLLSGFFINDRIPQYIMFHYLSLVAYPEAVLON 622

QY 579 EYFGLNFTCGSSNVSVTTNPMCAFTQCIQIEFTK 612

Db 623 EF-----SDPEECVRGVQLEFDS 641

RESULT 15

QY 09ZU35 PRELIMINARY; PRU; 725 AA.

AC 09ZU35

DT 01-MAY-1999 (TREMBlrel. 10, Created)

DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Putative ABC transporter.

GN ATG01320.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.

OC NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-CV. COLUMBIA;

RX MEDLINE-20083487; Pubmed-10617197;

RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,

RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,

RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,

RA Cronin L.A., Shen M., Vanden S.E., Umeyam L., Tallon L.J., Gill J.E.,

RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,

RA Salzbarg S.L., Fraser C.M., Venter J.C.;

RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis

thaliana.";

RT thaliana. 2;

RU Nature 402:761-768(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-CV. COLUMBIA;

RA Lin X.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

CC 1--SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.

DR EMBL: AC006200; AAD14532.1;

DR InterPro: IPR003593; AAA_ATPase.

DR InterPro: IPR003439; ABC_transporter.

DR Pfam: PF00005; ABC_tran; 1.

DR ProDom: PD000006; ABC_transporter; 1.

DR SMART: SM00382; AAA; 1.

DR PROSITE: PS00211; ABC_TRANSPORTER; 1.

DR ATP-binding; Transport.

SO KW

Query Match 725 AA; 7082556FE355307 CRC64;

Best Local Similarity 18.4%; Score 610.5; DB 10; Length 725;

Matches 166; Conservative 111; Mismatches 244; Indels 37; Gaps 11;

QY 51 VRP-----WMDITSC-----ROOQTRQILKDVSLVSGQIICIGSSGSKTTLLDMSGR 102

Db 65 IRVYTRMKNRITTSLSKSKSVRLKKNVSGAKGRILALNGPSSGSKTTLLNVLAV 124

QY 103 LGRAGT--FLGEYVNGRALRRQFODCSYVLSQPTLLSLVRETLHTTALLAIRR-G 159

Db 125 LSLSPRLHLSGLEVNGKRPSSKAYK--LAFVQEDLFSQTLVRETLSPFAALQLEIS 182

```

OY 160 NGSGOQKVEAVMABLSLSHVADRILGNTSLGISTIGERRRYSIAQLODDPKVMLFDEP 219
Db 183 SAEERDEYVNNLLMLGLVSCSDCVGDAKRVGISGGEKKRLSLACELASPSVIADEP 242
OY 220 TTGLDCAPIANOJVLVLLVLELARRNRRLTYVLIHOPRSLSFOLFPKIALISGELFEGTPA- 278
Db 243 TTGDJAPAEKRMEMFLQTLQADGHHYVICSIRHPRGSVYAKFDDIYLLTECTLYIAG-PMG 301
OY 279 -EMLDFNDCGIPCEPHSNPFEDYMDLNSVPTOSKEREIETSKRVQMIESAY--KSAIC 335
Db 302 KEPLTYFENFGELCEPHNVPAEFLDILLISVDSSSETVYSOSKRVHALVDAFSQRSSSVL 361
OY 336 HKTLKINRMKHLKTLPMVPFKTKDPSGVFSKGLYLRV-----TRNLYRNKLAYI 387
Db 362 YATPLSMKEERKNGRRPRKALVETDGMKROFPLLKKRAMMAQASDGEPTNVRARMSVA 421
OY 388 TRLLONLIMGLFLEFVLVRVNSVLKGAIDRVRGLLYOFVGATPYTGMLANVLFPEVLA 447
Db 422 SA-----VIFGSYFWRMGKSQTSIIDRMGLQVAAINRMALLKTKGVFPERA 471
OY 448 VSDQESDGLYOKQWOMLALAYLHPFVSVAAMISSVYKTLGLHPEVARRGYFSAALL 507
Db 472 IYDNERSGSISLGYLLSKTIAELFPGCAPPLMGAVLYPPARLNPLTSLRSGKEGCIY 531
OY 508 APHLIGEFLTYVLGTYONPNIVNSVALLSINAGLVSGSEFLRNIOEMPIPKIISYTFE 567
Db 532 VESFEAASAMGLTVCAMVPSTEAAMAVGPSLMTVFLVFG-GYVYVNDNPIILIRWIPRAL 590
OY 568 OKYCEILVNVEEGLNF 585
Db 591 IRMAPOGLCINFEESGLKF 608

```

Search completed: June 11, 2003, 09:13:23
Job time : 43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 09:09:57 ; Search time 22 seconds
(Without alignments)
2844.705 Million cell updates/sec

File: us-09-989-981a-6
Perfect score: 3326
Sequence: 1 MCDLSSITPCGSMGLQVNRG.....PALVILGIVFKIRDLISR 651

Scoring table: BIOSGM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

1: plr1:
2: plr2:
3: plr3:
4: plr4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	649.5	19.5	725	2 T47652	ABC transporter-11
2	627	18.9	1294	2 S77690	Probable membrane
3	621	18.7	1049	1 S19421	ATP-dependent perm
4	614	18.5	720	2 T47648	ABC transporter-11
5	610.5	18.4	725	2 C84423	Probable ABC trans
6	608	18.3	708	2 T47650	ABC transporter-11
7	602.5	18.1	687	1 FYEFW	white protein - fr
8	597	17.9	590	2 B96573	protein F12M16.17
9	595.5	17.9	658	2 T31958	hypothetical prote
10	591.5	17.8	740	1 T02567	Probable ATP-bindi
11	590.5	17.8	638	2 G02068	Probable ABC trans
12	589.5	17.7	646	2 C86441	white hemolys - hu
13	584	17.6	755	2 G84791	Probable ABC trans
14	583.5	17.5	687	2 D96553	hypothetical prote
15	578.5	17.4	646	2 JC7777	ATP binding caset
16	575.5	17.3	649	2 A84509	Probable ABC trans
17	572	17.2	739	2 T45891	ABC transporter-11
18	571.5	17.2	609	2 E96742	Probable ABC trans
19	564	17.0	662	2 T47649	ABC transporter-11
20	554	16.7	559	2 B89474	protein C05D10.3
21	552.5	16.6	635	2 T08334	hypothetical prote
22	545.5	16.4	608	2 T34391	hypothetical prote
23	540	16.2	678	2 H96552	hypothetical prote
24	539	15.9	577	2 T04429	ABC-type transport
25	529.5	15.9	633	2 T19189	hypothetical prote
26	527.5	15.9	659	2 E86313	hypothetical prote
27	517.5	15.6	610	2 T19333	hypothetical prote
28	517.5	15.6	639	2 G88839	protein C10C6.5
29	514.5	15.5	705	2 D84680	probable ABC trans

30	504	15.2	695	2 T21109	hypothetical prote
31	497	14.9	547	2 T31543	hypothetical prote
32	488	14.7	737	2 T46101	ABC transporter-11
33	464	14.0	1450	2 A84780	Probable ABC trans
34	456.5	13.7	1426	2 T30567	ATP-binding caset
35	454.5	13.7	675	1 FYEFB	brown protein - fr
36	435	13.1	668	2 S55023	brown protein - fr
37	429	12.9	1469	2 H96622	Probable ABC trans
38	425.5	12.8	1443	2 T02431	Probable ABC trans
39	425	12.8	1451	2 B86286	F9L1.15 protein -
40	421	12.7	1450	2 T45888	ABC transporter-11
41	420	12.6	1435	2 D96693	ABC protein putative A
42	418.5	12.6	1420	2 T02644	ABC-type transport
43	417.5	12.6	1619	2 T30541	ABC1 transport pro
44	413	12.4	1413	2 G84790	Probable ABC trans
45	410	12.3	1177	1 D71416	probable PDR5-like

ALIGNMENTS

RESULT 1					
T47652					
ABC transporter-like protein - Arabidopsis thaliana					
N:Alternate names: protein T26112.10					
C:Species: Arabidopsis thaliana (mouse-ear cress)					
C>Date: 20-Apr-2000 #sequence, revision 20-Apr-2000 #text, change 19-May-2000					
C:Accession: T47652					
R:Monfort, A.; Casacuberta, E.; Pulgomech, P.; Mewes, H.W.; Lemcke, K.; Mayer, submitted to the Protein Sequence Database, February 2000					
A:Reference number: Z24471					
A:Accession: T47652					
A>Status: Preliminary					
A:Molecule type: DNA					
A:Residues: 1-725 <MON>					
A:Cross-references: EMBL:AL132954					
A:Experimental source: cultivar Columbia; BAC clone T26112					
C:Genetics:					
A:Map position: 3					
A:Note: T26112.10					
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12M16.1;					
Query Match					
Best Local Similarity 29.4%; Pred. No. 1e-40; Length 725;					
Matches 182; Conservative 124; Mismatches 246; Indels 67; Gaps 15;					
QY	33	PEPSSIGILHASYSVSHRVRPMDITSCROOMPFOILKDYSLVESGOIMCIIGSSGSK	92		
DB	68	PVRYVNLNFNNLOVDVTLRR-----FGSSRONGKVTLLDDVSGEASGDILAAGSAGK	123		
QY	93	TTLLDANGSLGRLAGFFLGEVYNG-RALREDFQDFSVYLAOSDILLSLYVRETLAYT	151		
DB	124	STLLDLAGRAVAE-GSLRGSVTLNGEVRSLRLKVASAVYMODDLEFPMVKEITLMAFA	182		
QY	152	ALLAIRRG-NPGSPQKVEVAEMLSHVADRLIGVSLGSGTSGRRRVSTAQLOLD	210		
DB	183	SEERLPSLSKSKMEVEVALIDQLGRNANANVIGEGHGVSGGRRRRVSGIDIDHD	242		
QY	211	PKVMEDEPTGLDCMTANOIVLVLEARRNRIVVTHQPSLEFOLDKTAISFGE	270		
DB	243	PVLFDEPTSGIDSTNAFVVOYLKRIAGSGSVIASHQPARVELDKLITLSRGK	302		
QY	271	LIFCGPAENLDEFNDGCPSPHSNPFDEYMDLTSVDTOSKREIETS-----	319		
DB	303	SVFNGSPASLPGEFSDGRPIPEKENTISEFALDV-----RELGSNEGKALVDEN	354		
QY	320	-----KRYOMIEAAYK-----KSAIGHKTL--KNIRMAHLKTLPAVPRPTKD	360		
DB	355	EKWQOKNISLQSPQTNKLDORSLSLKAINASVSRGLVSGSSRSNPTSMETVSSYA	414		
QY	361	SPGVFSKLGVLRRVRLNKLAVITRLQNLINGLFLFVYLVRASVYKLGATODRV	420		
DB	415	NPSLFETF-ILAKRYKKNWIRMPELYGTIRATVAVTGC-TLATVYWKLDHPTPGA-OERL	471		

QY 421 GLLTQVVGATPTTGMNANVLPVLAASVDSODGLYQKQMDLAVALHLPFSVATM 480
 Db 472 -FLFAVVPPTMPCCLDNPVPEIOENYIFLRETHNAVTSYVISHLSVLPOLLAPSL 530
 QY 481 IFSSVCYWTIGLHPEVARGFSANLHLIGEFLVLVLGIVQNPVNS-VALLSI 539
 Db 531 VSAITFTVGLSGLEGVFTCLLYASFGSSVVTETISGV--PNIMLCWYSITPL 588
 QY 540 AGVLVSGFLRNIOEPIPKIISYFTFOKCSSEILVNEFGLNFTCCSSNVSVTNPM 599
 Db 589 AVCLLSGFYVRDRIRPFYTWPHYSILAKYEVANLINEF-----DDPS 633
 QY 600 CAPTQGIQPIETKTCGATS 618
 Db 634 RCFVRGVQVFDSTLLGVS 652

RESULT 2

Probable membrane protein YOL075c - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein O1125; hypothetical protein O1130; hypothetical
 C:Species: Saccharomyces cerevisiae
 C:Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 19-Apr-2002
 R:Accession: S77690; S66767; S66768
 R:Alexander, D.; Katsoulou, C.; Tzermia, M.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S66756
 A:Accession: S77690
 A:Molecule type: DNA
 A:Residues: 1-1294 <ALE>
 A:Cross-references: EMBL:Z74816; MIPS:YOL075c
 A:Note: this is a revision to the sequence from reference S66756
 A:Accession: S66767
 A:Molecule type: DNA
 A:Residues: 1-179, 'TTRGVFLYKRED' <ALM>
 A:Cross-references: EMBL:Z74816
 A:Experimental source: strain S288C
 A:Note: this sequence has been revised in reference S77690
 A:Accession: S66768
 A:Molecule type: DNA
 A:Residues: 200-1294 <ALP>
 A:Cross-references: EMBL:Z74817
 A:Experimental source: strain S288C
 A:Note: this sequence has been revised in reference S77690
 C:Genetics:
 A:Note: this was assumed to be the complete sequence of protein YOL075c
 C:References: SGD:S0005435
 A:Map position: 15L
 A:Note: YOL075c
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
 C:Keywords: ATP; nucleotide binding; P-loop; transmembrane protein
 F:45-263/Domain: ATP-binding cassette homology <ABC1>
 F:62-69/Region: nucleotide-binding motif A (P-loop)
 F:376-392/Domain: transmembrane #status predicted <TM1>
 F:469-485/Domain: transmembrane #status predicted <TM2>
 F:496-512/Domain: transmembrane #status predicted <TM3>
 F:606-622/Domain: transmembrane #status predicted <TM4>
 F:727-734/Region: nucleotide-binding motif A (P-loop)
 F:1042-1058/Domain: transmembrane #status predicted <TM5>
 F:1155-1141/Domain: transmembrane #status predicted <TM6>
 F:1177-1193/Domain: transmembrane #status predicted <TM7>
 F:1269-1285/Domain: transmembrane #status predicted <TM8>

Query Match

Best Local Similarity 18.9%; Score 627; DB 2; Length 1294;
 Matches 181; Conservative 106; Mismatches 228; Indels 56; Gaps 19;

QY 65 TROLIKDVLVSGVSGGKTTLLDMSGSLGRAGFL-----GEYVNRRA 119
 Db 706 TKELQSVNALPFGKMINALMGPSGSKSLNLISGHL-KSYFAKFDISGSIIMPDIQ 764

QY 120 LRREPODFCSYVLASDT-ILSLVRETLHYTALLAIRGNPGSPORKVEAVMAELSS 178
 Db 765 VSEIMKKNVCSTYSODDHLALALYKELTKAALRLHLTEAEMETDNLISGLK 824
 QY 179 HVADRLIGNSLAGISTGERRRVSIAAQLDPPKWLFEDEPTGLDCMTANOIVLVLEL 238
 Db 825 HCENNIKNEFKVIGSGEKRRRTMGVQLNDPILLDPEPTSGIDSLFSATILELTL 884
 QY 239 AR-RNRIVVLTHORSELEFQFDKAIIS-FGELFGSTPAMLDFFMDGCPCEHSN 296
 Db 885 CREQKQTTITTHQPSSELEFKRFGVLLAKSGRAFNQSPDEMTAYFELGYNCPSPFN 944
 QY 297 PPDFYMDLTSVDQSEKEREIETSKRQMIESAICHTKLNIERKHLKTLPMVFE 356
 Db 945 VADFFLDLSVTQNEQNEISSRAVERKILSMK-----ANNDR-ESLSPPISEK 994
 QY 357 KTRDSEGVFSKGLVLRRTNLY-----RNKLAVTRILLQIMGLFL 401
 Db 995 QOYSGSEPFTESEYVRK-PANLVLAYIYVNRQPTTTRSPDSLMARIAQIPGLCVIRA 1053
 QY 402 FFLVAVRSNVLKGAIQDRVGLLYQVGAATPYGMNAVNLFPVLAASVDSODGLYQRM 461
 Db 1054 LEFAVVKHNYT--SISNRLGLAQEST-ALYFGMLGNLACYPTERDYFEETNDNYGIA 1110
 QY 462 QMMLAYALHVPFSVATMIFSSVCYWTIGLHPEVARRCYFSALLAPHLI---GEFLTL 518
 Db 1111 PFTLAVMTLEPLSLASVLAFTVLACGL-PRTA-GNEFAFYCSFIVCCGERLGI 1167
 QY 519 VLLGIQVQNP-IVNSVVALISAGVLVSGFLRNIOEPIPKIISYFTFOKCSSEILV 577
 Db 1168 MNTFEPFRPGFVNVNCSIIISIGTQMSGLMSL---GMSRYLKGNYNLNPGVYTSMTIIN 1223
 QY 578 NEFTG-LNFTG--GSSNVSVTTNPGCAFTQG 605
 Db 1224 FAFPGNLKLTCEDEGCKNSDGT---CEPANG 1250

RESULT 3

ATP-dependent permease ADP1 precursor - yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein YCR011c; protein YCR105
 C:Species: Saccharomyces cerevisiae
 C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
 R:Goffeau, A.; Purnelle, B.; Skala, J.
 submitted to the Protein Sequence Database, March 1992
 A:Reference number: S19421
 A:Accession: S19421
 A:Molecule type: DNA
 A:Residues: 1-1049 <GOF>
 A:Cross-references: EMBL:X59720; NID:q1907116; PIDN:CAA42328.1; PID:q1907154; GSPDB:
 R:Purnelle, B.; Skala, J.; Goffeau, A.
 Yeast 7, 867-872, 1991
 A:Title: The product of the YCR105 gene located on the chromosome III from Saccharo
 A:Reference number: S40914; MUID:92160395; PMID:1789009
 A:Accession: S40914
 A:Status: not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-1049 <PDR>
 R:Skala, J.; Purnelle, B.; Goffeau, A.
 Yeast 8, 409-417, 1992
 A:Title: The complete sequence of a 10.8 kb segment distal of SUR2 on the right arm
 K genes.
 A:Reference number: S25353; MUID:92327849; PMID:1626432
 A:Contents: annotation
 C:Genetics:
 A:Gene: SCD:ADP1; MIPS:YCR011c
 A:Cross-references: SGD:S0000604; MIPS:YCR011c
 A:Map position: 3R
 C:Superfamily: ATP-dependent permease ADP1; ATP-binding cassette homology
 C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop; transmembrane protein
 F:1-35/Domain: signal sequence #status predicted <SIG>

F:26-1049/Product: ATP-dependent permease ADP1 #status predicted <MAY>
 F:26-324/Domain: extracellular #status predicted <EXT>
 F:323-341/Domain: transmembrane #status predicted <TM1>
 F:406-607/Domain: ATP-binding cassette homology <ABC>
 F:423-430/Region: nucleotide-binding motif A (P-loop)
 F:550-557/Region: nucleotide-binding motif B
 F:794-810/Domain: transmembrane #status predicted <TM2>
 F:829-845/Domain: transmembrane #status predicted <TM3>
 F:878-894/Domain: transmembrane #status predicted <TM4>
 F:909-925/Domain: transmembrane #status predicted <TM5>
 F:938-954/Domain: transmembrane #status predicted <TM6>
 F:1025-1041/Domain: transmembrane #status predicted <TM7>
 F:50-114,165,221/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:429/Binding site: ATP (Lys) #status predicted

Query Match 18.7% Score 621; DB 1; Length 1049;
 Best Local Similarity 28.6% Pred. No. 2.3e-38;

Matches 196; Conservative 111; Mismatches 223; Indels 156; Gaps 22;

Db 68 ILKDVSLVBSGQIMCIISSGSGKTTLLDAMSGRLGRAGTFLEGVYVNGRRLRREOFOD 127
 405 VLNEISGIVRGQITLIMGSGGACKTLLDILAMK-RKTHVSGSIVNGISMOKRGSFK 463
 Oy 128 CFSVYLOSDFLLSLTYRETLHTALLAIRGNPFSOK--VEAVMAELSLSHVADRL 184
 Db 464 TICFVDDDELPLTYEFTVLSALRLPRAL--SEAKKARYKYLEIRIDIRI 521
 Oy 185 IGNTSLGISTGERRRSIAAOLDDPKVLFDEPTGIDCMANOIVLLVELAR-RNR 243
 Db 522 IGNEFDGIGGERRRSIACELTSPLEFLDEPTGLDLSNANNVIECLVRLSSOYNR 581
 Oy 244 IVVLTIHQPSSELEQLDKITALLISFGLIFCGTPAEMLDEFNDGCGPEHSNPFEDMD 303
 Db 582 TLVLSIHQPSNIFYLEDKLVLSKGMVYSGNAKKYSEFLRNGYICPDVNIADLID 641
 Oy 304 LT-SVDTQSKEREI----- 316
 Db 642 ITPEAGPQGRRRIRNISLEAGTDNDIDNTIHQTFTSSDGTTOREMAHLAHRDEIR 701
 Oy 317 -----ETSKRVOMIESAYKSAICHKTKNIEM----- 345
 Db 702 SLRDEEDVSTGCRGKATEIDNTKILHDKTYDSVYAEISOEIEVLSEGDSESVLN 761
 Oy 346 KHLKTLPMVPFKTKDSPGVSKGLVLLRYTRNLVRKLAIVTRILLQNLIMGLFLEFVL 405
 Db 762 GDLPT-----GQSSAGFLOQLSILNSRFKNRYRNPRIKLLGYLLTILLSLEFLGLTY 814
 Oy 406 RYNSNVLKGAIDORGLTY---OPVGTPTTGLMAVNLFPVLAASVDSQDGLYQKWQ 462
 Db 815 NV-SNDISG-FQNRMGLEFFILTYFVFVFTGL-----SSFALERIIIFIKERSNNYSP- 866
 Oy 463 MMLAYAL-----HYLPESVAVATMIFSSVCYTGLIHEVARFGFSALLAPHLIGELT 517
 Db 867 --LAIYISKINSEVPLRAVPPILSLITYPMGLNKNDAF-FKICIGILLENLGLISLE 923
 Oy 518 LVLLGIV---QNPNIYSVALISIAVGLVSGFLRNIOEM-IPKTIISYTFQKRCSE 573
 Db 924 ILTIGLIFEDLNISILSVLVL---GSLFSGLFITKNTITNAVFYLANFESFYAYE 980
 Oy 574 ILVYNEF-----YGLNFTGSSSVSTTNPMACTGOGIOIEKTCGCAFSRTMNF 624
 Db 981 SLILNEVKTLMLEKRYGLNI-----EYGCATILLSTFGF 1014
 Oy 625 LILYSEFIPALVIGI--VVEKIRDL 648
 Db 1015 -VQNLVPEIKITALLFNVVFLMGYL 1039

RESULT 4
 T47648
 ABC transporter-like protein - Arabidopsis thaliana
 N:Alternate names: protein T15C9, 80
 C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
 C:Accession: T47648
 R:Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
 submitted to the Protein Sequence Database, April 2000
 A:Reference number: 224470
 A:Accession: T47648
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-720 <MEM>
 A:Cross-references: EMBL:AL132970
 A:Experimental source: cultivar Columbia; BAC clone T15C9
 C:Genetics:
 A:Map position: 3
 A:Note: T15C9, 80
 C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1;

Query Match 18.5% Score 614; DB 2; Length 720;
 Best Local Similarity 28.7% Pred. No. 4.6e-38;

Matches 182; Conservative 122; Mismatches 254; Indels 76; Gaps 17;

Oy 23 SLEEG--APATAPEPHSLGILHASYVS-----HVRPMWDITSCROQTOILKDV 73
 Db 40 SSLDGDNDHLMRPVPEVLSFNNTLYNVSVRRKIDFHLVPMRRTSFSG--TKTLNDIS 96
 Oy 74 LYVESGQIMCIISSGSGKTTLLDAMSGRLGRAGTFLEGVYVNGRRLRREOFOD 133
 Db 97 GETRDGELTAVLCASSGSGKTTLLDILANRLK-GLKGYTLNLEALQSLMKLYISAVM 155
 Oy 134 QSDTLSSLTVRETLHTALLAIRGNPFSOK-KVAVMAELSLSHVADRLIGNYSILG 192
 Db 156 ODDLFPMLVVEETLMAPELRPLRSPLKSKKRLVQALLDQGLIRNAATITIDEBHRG 215
 Oy 193 ISTGERRRVIAAOLDDPKVLFDEPTGIDCMANOIVLLVELARRNIVVLTIHQ 252
 Db 216 ISGERRRVISIGDIIDHPVLEFLDEPTGLDSTSAVMYKVLRIEESGIIIMSITHQ 275
 Oy 253 RSELFQLEDKITALLISFGLIFCGTPAEMLDEFNDGCGPEHSNPFEDMDLS----- 306
 Db 276 SHRYLSLDLRLIFLSRQHYFSGSPASLPSEFAGFPDENQTEFALDLRLREGSA 335
 Oy 307 -----VDTSKEREIETSKRVOMIESAYKSAICHKTKL-----NIEMKHLK----- 349
 Db 336 GGTGRGLVEFKNKQEMKQNSPQTLPPASPNP--NLTKELIASASISRGLVSGGGGS 393
 Oy 350 -----TLPMVPFKTKDSPGVSKGLVLLRYTRNLVRKLAIVTRILLQNLIMGLFEL 401
 Db 394 SVINHGGGLAVAFANP-----FWIEIKTLTRRSILNSRQPELGLRLAVIYTG-FIL 448
 Oy 402 FVYLRYNSVNLKGAIDORGLTYOFVGTPTTGLMAVNLFPVLAASVDSQDGLYQKW 461
 Db 449 ATYFRLNDNSPKG-VOERLG-FFAFANSTMYTCADLPFLDERVYFMTAVNARRS 506
 Oy 462 OMMLAYALHYLPESVAVATMIFSSVCYTGLIHEVARFGFSALLAPHLIGELT 521
 Db 507 SYVLSHAIVPEPSLIFLSLAPAVTTFAVGLLEGIMGLFELCYLLIASFSFGSFFVFLS 566
 Oy 522 GIYQNRIV---NSVALISIAVGLVSGFLRNIOEMIPKTIISYTFQKRCSEILVN 578
 Db 567 GVV--PRVMLGYTIVAIL--AYELLESGFEINDRIPQYIMFHYLSLVKYPYEAIVLN 622
 Oy 579 EYGLNFTGSSSVSTTNPMACTGOGIOIEKTE 612
 Db 623 EF-----SDPTECFVKGVLFDMS 641

RESULT 5
 C84423
 Probable ABC transporter (imported) - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: C84423
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.T.; Town, C.D.; Fujii,
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Taii

enus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A:Reference number: A84420; MID:20083487; PMID:10617197

A:Accession: C84423

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-725 <STO>

A:Cross-references: GB:AE002093; NID:94262339; PIDN:AAD14532.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g01320

A:Map position: 2

Query Match 18.4%; Score 610.5; DB 2; Length 725;
Best Local Similarity 29.7%; Pred. No. 8.6e-38;

Matches 166; Conservative 111; Mismatches 244; Indels 37; Gaps 11;

```

OY 51 VRP-----WMDITSC-----RQMTROLKQVSLVSGQIMCIGSSGSKTLLDMSGR 102
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 65 IRPVTIRMNITCSLSDSKSVRFILKNVSGEAKGRLLAIMGPSGKTLLANVLAGO 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 103 IGRAGT--FLGEYVNGRALLRREPQDCSYVLOSOTLLSLVRETLATYLLAARR-G 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 125 LSLSPRLHSLGLEPNGKSSSKAYK--LAFVRQEDLFFSOLVRETLSPFAELDLPETS 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 160 NGSFQKRYEAVMAELSLSHVADRILIGNSLGISTGERRRYSIAQLIDPKVMLFDEP 219
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 183 SAEDEYVNNLLKGLGVASCADSCGDAKVGISGGEKKRLSLACELIASPSVIFADEP 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 220 TTGLDQMTANQIVLVELARRNRIVYLTHOPRSEFLQEPKIALISFGLIFCGTPA- 278
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 243 TTGLDAFOAEKMETLQKLAODGHVYICSIHOPRGSVYAKFDIYLLFEGTIVYAG-PAG 301
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 279 -EMLDFNDGYPCEPHSNPFEDYMDLTSVDOSKREIETSKRVOMIESAY--KKSATC 335
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 302 KEPLVYFGFGLCEHVPAPFLADLVSDYSSSEYVSSQRRHALDASQSSSVL 361
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 336 HKTINIERMKHLKTLPMVPEYTKDSPGVFSKLGVLRRV-----TRNLVNRKLAVI 387
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 362 YATPLSMKEETKNGMPRRKAIYERTDGMWRQPFILKRAMQASNDQTNVBARMSVA 421
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 388 TLLQNLIMGLFLFLFVLRVRSNYLKGAIQDRVGLLYQVGAFTPTGMLNANVLFPLA 447
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 422 SA-----VIRGSYFMRMGKSQTSIQDRMGLQVVAINTMAALTKYGVPEKERA 471
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 448 VSDQSDGLYOKMOMLALAYLHVLPFSVATMIFSSVYWTGLHPEVARRGYSAALL 507
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 472 IYDRERSKGSYSLGPIILSKTIAETIPICAFPLMGAVLYPMARLNPILSRGKFCGIVT 531
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 508 APHLIGELTLVLLIGIVONPNIVNSVVALLSIAGVLYSGSGLRNIOEMPPIPKIISYTF 567
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 533 VESFASAMGLTVGAMVPESTEAMAVGPELMTVFIVFG-GYVYVNDNPIIFRMIPRASL 590
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 568 OKYCEIILVNEFYGLNF 585
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 591 IRMAFOGLCTINEFSGLKF 608
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 6

T47650
ABC transporter-like protein - *Arabidopsis thaliana*
W:Alternate names: protein T15C9.110
C:Species: *Arabidopsis thaliana* (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence-revision 20-Apr-2000 #text-change 19-May-2000
C:Accession: T47650
R:Reviews, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24470
A:Accession: T47650
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-708 <HEW>
A:Cross-references: EMBL:AL132970

A:Experimental source: cultivar Columbia; BAC clone T15C9
C:Genetics:
A:Map position: 3
A:Note: T15C9.110
C:Superfamily: *Arabidopsis thaliana* probable ATP-binding cassette protein P12L6.1;

Query Match 18.3%; Score 608; DB 2; Length 708;
Best Local Similarity 27.5%; Pred. No. 1.3e-37;

Matches 111; Conservative 136; Mismatches 256; Indels 58; Gaps 15;

```

OY 18 NNGSGSLGADP--TAPPHSLGILHASYSHVSRVPMWDTSCROMTROLKQVSLY 75
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 41 NAPTQHILADPAEATRSVPFLSFNNLSYNNVLR--FDSRRKATAVKTLDDITSE 98
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 76 VESGQIMCIGSSGSKTLLDMSGRRLRAGTFLCEYVNG-RALRRQFDDCESTVQ 134
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 99 ARDGETLAVIGSGGAKSTLIDALGRVAE-PSLKTQVNLNGEKVQLSRLKIVASVYMQ 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 135 SDTLLSLVRETLHTALLAIRGNGSFO-KRYAVAAELSLSHVADRILIGNVSLGI 193
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 158 DDLPPMLTVKETLFASEFRLPRSLPKSKMEHREVELIDQLGRNAADTVIGDCHRCV 217
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 194 STGERRRYSIAQLIDPKVMLFDEPTGIDCMTANQIVLVELARRNRIVYLTHOPR 253
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 218 SCGERRRYSIGIDIHDPILFLDEPTSLDSTNFMVVOYLKRIQSGSVYIMSHOPS 277
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 254 SELFQELKIALISGELIFCGTPAEMLDFNDGYPCEPHSNPFEDYMDL-----TGV 307
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 278 ARITGLDLRLILSHKSVFNGSPVSLPSFSSFGPRIPKEKENTIEFALDVIRESGSE 337
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 308 DT-----OSKREIETSKRVOME-----SAYKSAICHKTINKIERMKHLKTL 352
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 338 GTRDLVERNEKQONQATATQOSKSLKALIASYSRKGLVSGSGAPISMETVSYA 397
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 353 MVEPFTKOSPGEFSKLGVLRRVNRNLVNRKLAVITRRLQNLIMGLFLFLFVLRVSNYL 412
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 398 NPP-----LAEFLIAKRYIKMINTPELIGRICTGVVGTLLATVVRLL-DNRP 447
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 413 KGAIDRQGLYQFGAPFTYTGMLNANVLPFLRAVSDQSDGLYOKMOMLALAYL 472
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 448 RGA-DEKRG-PFAFGSTVFYCCADNIPFQIDERYIFLSETHNMYRRSSSYISHALVSL 505
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 473 PFSVATMIFSSVYWTGLHPEVARRGYSAALLAPHLIGELTLVLLIGIVONPNIVNS 532
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 506 PQLMALSTAFATTTWVGLSGSLGSEFYICLIITAAWSSGIYTFISGL--PVMMS 563
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 533 -VALLSIAGVLYSGSGLRNIOEMPPIPKIISYTFQKYSILVNEFYGLNFTGSSN 591
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 564 YWVTIAYLSYCLLGGFYINRDRIDPLYMIMFYISILKPYEAVLINER----- 612
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 592 VSVTNPMCAFTQGIQFIETK 612
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 613 ----DPSRCFVGQVDFDGT 629
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 7

FFFW

white protein - fruit fly (*Drosophila melanogaster*)

C:Species: *Drosophila melanogaster*

C:Date: 31-Dec-1990 #sequence-revision 17-Feb-1995 #text-change 19-Jan-2001

C:Accession: S08635; S07263; S10240

R:Pepling, M.; Mount, S.M.

Nucleic Acids Res. 18, 1633, 1990

A:Title: Sequence of a cDNA from the *Drosophila melanogaster* white gene.

A:Reference number: S08635; MID:90221897; PMID:109311

A:Accession: S08635

A:Molecule type: mRNA

A:Residues: 1-687 <PEP>

A:Cross-references: EMBL:X51749; NID:98825; PIDN:CA36038.1; PID:98826

R:O'Hare, K.; Murphy, C.; Lewis, R.; Rudin, G.M.

J. Mol. Biol. 180, 437-455, 1984

A:Title: DNA sequence of the white locus of *Drosophila melanogaster*.

A:Reference number: S07263; MID:85134865; PMID:6084717

Db 646 GLVALKIR 653

LEONOVHE-----SAINSAICHNITUNIERK----- 346

|| : | : || : | :

LEONOVHIE-----SAINSAICHNITUNIERK----- 346
|| :| : | :: || : :

QY 347 ----HLKTLPMVPEFKTKDSQVCSKGLVLLRVTNRLVNRKLAIVITRLLONLIMGLEFLF 402
 DB 415 GATTNTTTLAVPAPFAPD-----MMIEIKTLKSRMLNSLRQPELFGIRIAGVITG-FILA 469
 QY 403 FVLAVRSNVLKGAIDORVGLLYQFVGATPYTGLNVLNVPVLRANSDDOSOGGLQKMQ 462
 DB 470 TVFWRLDNSPKG-VQERLIG-FFAFAMSTMEYTCADLVPVLOERYITFMRATAYNARRSS 527
 QY 463 MMLAVLHVLPESVATMIFSSVYWTLCGLPEVARGPSALLPHELIGEFLTVLLIG 522
 DB 528 YVLSHAIIVSPSLIFLSVAPAAATTYAVGLDGLTGLIFCLITLISFMSGSSFTPLSG 587
 QY 523 IVONPNIV---NSVALLSTAGLVGSGFLRNIOEMPIPEKIIISYTFPOKCEIIVVNE 579
 DB 588 VV--PSVMLGTYTIVAIL--AYPLFSGFFINRNRIIPDWIMFHYMSLVKYPEAVLQNE 643
 QY 580 FYGLN--FYCG-----SSNVSVTTPNMCATQG----- 605
 DB 644 FSDATKCFVAGVQIFDNTPLGELPEVMKLLGLTVKRSIGVITISSTCLTTCGSDILRQOG 703
 QY 606 -IOFIETPCGATSRFTMNFLLYSFIPALVILG 638
 DB 704 VVQLSKNCLFTITVAGFEFFRILIFY---TLLIG 734

RESULT 11

G02068
 White homolog - human
 C:Species: Homo sapiens (man)
 C>Date: 21-Dec-1996 #sequence, revision 06-Jun-1997 #text, change 02-Feb-2001
 C:Accession: G02068
 R:Crop, J.M.; Tiller, G.; Fletcher, J.A.; Lux, M.; Raab, E.; Goldenson, D.; Archilegas,
 submitted to the EMBL Data Library, August 1995
 A:Reference number: H00769
 A:Accession: G02068
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-638 <CRO>
 A:Cross-references: EMBL:U34919; NID:g1314276; PIDN:AAC51098.1; PID:g1314277
 C:Genetics:
 A:Gene: white
 C:Superfamily: fruit fly white protein; ATP-binding cassette homology
 C:Keywords: ATP; nucleotide binding; P-loop
 F:61-253/Domain: ATP-binding cassette homology <ABC>
 F:78-85/Region: nucleotide-binding motif A (P-loop)

Query Match 17.8%; Score 590.5; DB 2; Length 638;
 Best Local Similarity 26.3%; Pred. No. 2.3e-36;
 Matches 164; Conservative 142; Mismatches 266; Indels 51; Gaps 14;
 QY 44 SYSVSHRVPMMDITSCROQWROLIKDVSLVYESQIMCIGSSSGKTTLLDMSGRU 103
 DB 43 STSVBE--GPMW-----KKKGYKTLKIGISGFNSGELVAVIMGPSGAGSTLMNIIAGY- 94
 QY 104 GRAGTFLGEVYNGRRLRREOFQDFCFYVLOSITLLSLTVRETLHTVLLAIRGNPSS 163
 DB 95 -RETQMKGAVALINGLPDRLCFRKRYSCYIMQDKMLPHITVGEAMVASHLKIQEDGR 153
 QY 164 FOKKAYVAEELSHVADRLLIGNSLGISTGERRRYSIAQLODDPKVLMFDEPTGL 223
 DB 154 -REMYKEITLALGLSCA-----NTRTGLSGGQRRLAIALELVNPNVPMFDEPTSL 207
 QY 224 DCMFANOIVLVLELARRRIVVLTTHORSLEFOLFKIALISREBELFCGTPAMDF 283
 DB 208 DSASCFQVAVSLMKGLQAGRSITICTHOSAKLAFELFDLYVLSOGQCYRGRVNCVLY 267
 QY 284 FNDGCPSPHNSPFDYMDLTSDVTPQSKEREIETSKRVOMTESAKKSA-----I 334
 DB 268 LRDGLNCPYHNPAADFVAVEVASGEIGDQNSRLVAVRGMCDSDHKRLGDAEYNPPL 327
 QY 335 CKTKTKNTERMKHLTKLTPVAPKTKDSQV-----FSKGLVLLRVTNRLVNRKL 384

DB 328 WHRSEEVQKTKRLKL-----RKDSSMEGCHSFASCIUFOCLIFKRTFLSIMRDSV 381
 QY 385 AVITRLLONLIMGLEFLFVLRVRSNVLKGAIDORVGLLYQFVGATPYTGLNVLNVPV 444
 DB 382 LTHLRITSHIGLILGLTLYGIGNEAKK--VLSNSSFSSMLFELFALMPVLEPL 439
 QY 445 LRAVSDQSDGLYOKOMMAYALVHLVLPESVATMIFSSVYWTLCGLPEVARGPSA 504
 DB 440 EMGVLELHNMYSLKAYLLAKTMADVPQIPEPVAAGSYVWMSQSPDNAFVLPAA 499
 QY 505 ALAPHLIGEFLTVLLIGIYQNPVINSVALLSIAGVLVSGFLRNIOEMPIPEKIIISY 564
 DB 500 LGTMTSLVIAOSGL-LIGASTSIQVATFEVGPVTAIPVLLFSGFEVSFDITPYLQWMSY 558
 QY 565 FTQKCESELVVNEFYGLNFGTFGSSVSVSTTPNMCATQGIETCTGATSRFTMNF 624
 DB 559 ISTVRIGFEGVILS-ITGLD---REDLHCDIDETCHF-OKSEALIEIDVENAKYLD 612
 QY 625 LIYSEFIPALVILGIYV--FKIR 645
 DB 613 IVIGIFPISRLIAVPLRYKIR 635

RESULT 12

C86441
 probable ABC transporter (imported) - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Mar-2001 #sequence, revision 02-Mar-2001 #text, change 31-Mar-2001
 C:Accession: C86441
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; A
 Chn, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewa
 ansen, N.F.; Hughes, B.; Hultzer, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; K
 C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Mar
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Ta
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; M01D:21016719; PMID:11130712
 A:Accession: C86441
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-646 <SNO>
 A:Cross-references: GB:AE005172; NID:g11136734; PIDN:AAG31315.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1
 C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein f12L6.1;

Query Match 17.7%; Score 589.5; DB 2; Length 646;
 Best Local Similarity 29.9%; Pred. No. 2.8e-36;
 Matches 183; Conservative 112; Mismatches 246; Indels 71; Gaps 22;

QY 10 GGSM--GLQVNRSGS--SLEGAPATAPPEPISLITLASSVS-----HRYRPMMDITSCR 61
 DB 14 GGVNAGLDPMDSTQSKSVIAPFTITSQF---GLQSMPTITLKEVYKVK-LEQTSQCA 69
 QY 62 QQW---TROLKDVSLVYESQIMCIGSSSGKTTLLDMSGRGLRAGTFLGEVYNGR 118
 DB 70 GSWRSKEKTLINQITGMVCPGEFLAMGPSGSKTLLSALGRLSK--TFSGKYVNG- 126
 QY 119 ALRREFQOCF-----SYVLOSITLLSLTVRETLHTALLAIRGNPSPQK-----KYE 169
 DB 127 ----QPSGCCIRRRGFAVADVDVLPHTLVWETLFTALLRL-----PSSLTRDEKAHD 178
 QY 170 AVAAEELSHVADRLLIGNSLGISTGERRRYSIAQLODDPKVLMFDEPTGLDCMTAN 229
 DB 179 RVIAELGLNRCINSMIGGLFRLGISGEEKRRYSIGQEMLINSLLADEPTSGLDSTYAH 238
 QY 230 QIVVLLVLELARRRIVVLTTHORSLEFOLFKIALISREBELFCGTPAMDFDEPDGCV 289
 DB 239 RIYTTIKRLASGGRVVTTHOPSSRIYHMPKQVLLSGSPITTYGAASAVAEYFESSLGF 298

```

OY 290 PCPEHNPDPFYMDLTS---VDQSKERELETSKRVOMIESATK-----KSAICHKTL 339
DB 299 STSLTVNPDLDLLDNLGIPDPQKETSQEQKTVETLVSAEENKISTKRLKLNMASS 358
OY 340 KNIERAK-HUKTLPMVPFETKDSPGVSTGLVLRVETNL---VANKLAVITRLQNL 395
DB 359 HSEYETKAANKLKSQKCTT---WVQFTVLLQGVREERRESEFNKRLITQVISAVAL 414
OY 396 MCLFLLFVLRVSNVLKGAIDRVGLTYQVGAFTYQMLNANVLPYLRAVSDQSD 455
DB 415 GG--LLMW-----HFKSHIDRFTALLFFSVTFWGFPLYNANVTFPEKRLIKERS 466
OY 456 GLYQKQMLAYALAVLPFSVAVATMFSSVCTWTGLHPEVAFGYSALLAPHLIGEF 515
DB 467 GMYRLSSYFMANVGDLPLELALPTAFVFIYWMGGLKDPPTFFILSLVLYSVLAOG 526
OY 516 LFLVLLGIYQNPNIYNSVVALISGVLVSGFLRNIOEMPIPKI-ISTYFFQKCSFI 574
DB 527 LCLAGCALMNKQATTLASVTTLVFLNGYI---VQDIP--PFIYMLKLSYSTCYKL 582
OY 575 LVVNEFYGLNFT 586
DB 583 LL-----GIOYT 589

```

RESULT 13

```

G84791
Probable ABC transporter [imported] - Arabidopsis: thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: G84791
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;
Natus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: G84791
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-755 <STO>
A:Cross-references: GB:AE002093; NID:94056489; PIDN:ACG98055.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g37360
A:Map position: 2
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-B

```

```

Query Match 17.6%; Score 584; DB 2; Length 735;
Best Local Similarity 27.2%; Pred. No. 9e-36;
Matches 173; Conservative 128; Mismatches 253; Indels 82; Gaps 19;

OY 21 SSSLEGAPAT--AEPHSIGILHASYSVSHRVRPMDITSCROO-----TRQILNDV 72
DB 79 SFNSMASAPASISSPFVLSPTDLTYVKIO-KFNPPLACCRGNDSSVNTKILLNGI 137
OY 73 SLVYESGOIMCILLSSGSGKTTLLDAMSGRLGRACFTLGEYVYNGRALRREQFQDCFSY 132
DB 138 SGEAREGEMAVLIGASGSGKSTLLDALANRIAK--DSLNGSITINGEVLESSQKYSAYV 196
OY 133 LOSOTLLSLVRETLHTALLAIRGNPGSFQK-----VEAYMAELSLSHVADRLGN 187
DB 197 MODDLFPMVLVEETLMSAEFR---PRSLSKKKKARARVALDQGLSNAATVIGD 252
OY 188 YSLGISTGERRRYSIAQLDQPKVMLDEPTTGDCMTANOIVLVLLVELARRRRIYVL 247
DB 253 EGHNGVSGGERRRYSIGDITIDHPIILFDEPTSGLDSTSAVMYKVLQRIAGSGSIYIM 312
OY 248 TIHOPRSELFOLEFKIALISFGLIFCGTAPKMLDFNDGYPCEPHSNPDPFYMDLTSV 307
DB 313 SIHOPSRIMGLDQLIFLSKGNVTYSGSPTHLPQFSEKHPIDPENENKTEFADLTL-- 370
OY 308 DTQSKERELETSKRVOMIE-----SAVKSALCHKTLKNIERAKHLKTLP----- 352

```

```

DB 371 ----RELEYSTEGTKLVEFHKQAKAPSYNN---NKRNNTVSLKEALTAISRK 423
OY 353 -----MV- FKKDSPGVSKGLVLRTRNLVNRKLAIVITRLQNLIMGLF 399
DB 424 LVSGATNNSSNLTPEFQFANP-FWIEHIVIGKRALNSRQPELLGKRLGAVMTGII 482
OY 400 LTFEVLVRVSNVLKGAIDRVGLTYQVGAFTYQMLNANVLPYLRAVSDQSDGLQ 459
DB 483 LATMFNL-DNSKGA-QERLG-FFAFMASTTYTQAEIIPVLOERTYFMRETAYNAR 539
OY 460 KQOMLAVLHVLPFSVAVATMFSSVCTWTGLHPEVAFGYSALLAPHLIGEFITLY 519
DB 540 RSSYVLSQSIISIPALIVLASFAATFVANVGDGANGFEFFFTIILSFNAGSGFYTF 599
OY 520 LGIYQNPNIY---NSVALLSAGLVSGFLRNIOEMPIPKIISTYFFQKCSFI 576
DB 600 LSGVLT--PVMVGEFTYVAIL--AYELLSGFFISDRPIYVLMFHYLSLVKYPYEGVL 655
OY 577 VNEFYGLNFTCGSSNVSVTTPNCAFTQGIQFIETK 612
DB 656 QNEF-----QNPTRCARVQLEFNS 676

```

RESULT 14

```

D96553
Hypothetical protein F5D21.6 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: D96553
R:Theologis, A.; Ecker, J.R.; Palm, C.D.; Federspiel, N.A.; Kaul, S.; White, O.; Ali-
Chen, C.W.; Chung, B.K.; Comol, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar
ansen, N.F.; Hughes, B.; Huizar, L.
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Klu
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marz
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tal
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D96553
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-687 <STO>
A:Cross-references: GB:AE005173; NID:9100234349; PIDN:AMG12758.1; GSPDB:GN00141
C:Genetics:
A:Gene: F5D21.6
A:Map position: 1
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1;

```

```

Query Match 17.5%; Score 583.5; DB 2; Length 687;
Best Local Similarity 28.0%; Pred. No. 8.6e-36;
Matches 167; Conservative 117; Mismatches 239; Indels 73; Gaps 16;

OY 33 PEPHSIGILHASYSVSHRVRPMDITSCROO-----TRQILNDVSLVYESGOIMCILLSS 88
DB 13 PPPAPIC--KRAYLA-----WEDITVYIPNFGSGPTRLDGLGHPGKIMAGPS 64
OY 89 GSGKTTLLDAMSGRLGRACFTLGEYVYNGRALRREQFQDCFSYVLSQSTLLSLVRETL 148
DB 65 GSGKSTLLDLAGRLARVIMTGNLLNGKARLD--YGLVAYVQEDILMGTLLVRETI 122
OY 149 HTYLLAIRRGNPGSFQK-----VEAYMAELSLSHVADRLGNISLGIETGERRRYSI 203
DB 123 TYSALTRL-----SSDLTKEEVNDIYEGTIIELGLDCAADRYIGMHSNGVSGGERRRYSV 178
OY 204 AAQLDQPKVMLDEPTTGDCMTANOIVLVLLVELAR-RNRIYVLTTHOPSELFQLEDK 262
DB 179 ALEILTRQIILFDEPTSGLDSTSAVMYKVLQRIAGSGSIYHOPSSVYFALDD 238
OY 263 TAILSPGELIFCGTAPKMLDFNDGYPCEPHSNPDPFYMDLTSVDQSKERELETSKRV 322
DB 239 LFLLSGGEYVYFGSKFVVEFAAGFPCCPKRRNPSDHLFRLCINSDDIYATILKGSORI 298

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 09:12:22 ; Search time 25 Seconds

(without alignments)
2688.377 Million cell updates/sec

Title: US-09-989-981a-6

Perfect score: 3326

Sequence: 1 MGDLSTLFGSGMGLOVNRG.....PALVILGIVFKIRDLISR 651

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 392085 seqs, 103240269 residues

Total number of hits satisfying chosen parameters: 392085

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubppa/US08_NEM_PUB.pep:*
- 2: /cgn2_6/ptodata/1/pubppa/ECT_NEM_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubppa/US06_NEM_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubppa/US07_NEM_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubppa/ECTUS_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubppa/US09_NEM_PUB.pep:*
- 10: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubppa/US10_NEM_PUB.pep:*
- 12: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/1/pubppa/US60_NEM_PUB.pep:*
- 14: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3326	100.0	651	US-10-090-455-6	Sequence 6, Appl1
2	3326	100.0	651	US-09-989-981A-6	Sequence 6, Appl1
3	3326	100.0	651	US-09-837-992-3	Sequence 3, Appl1
4	2744.5	82.5	652	US-09-989-981A-2	Sequence 2, Appl1
5	2744.5	82.5	652	US-09-837-992-1	Sequence 1, Appl1
6	697	21.0	672	US-09-989-981A-4	Sequence 4, Appl1
7	697	21.0	673	US-10-090-455-7	Sequence 7, Appl1
8	697	21.0	673	US-09-989-981A-8	Sequence 8, Appl1
9	682.5	20.5	655	US-09-961-086-1	Sequence 1, Appl1
10	680.5	20.5	655	US-09-981-353-35	Sequence 35, Appl1
11	680.5	20.5	655	US-10-120-687-61	Sequence 61, Appl1
12	674.5	20.3	655	US-10-090-455-5	Sequence 5, Appl1
13	674.5	20.3	655	US-09-866-866A-10	Sequence 10, Appl1
14	672.5	20.2	655	US-09-866-866A-27	Sequence 27, Appl1
15	660	19.8	657	US-09-866-866A-14	Sequence 14, Appl1
16	602.5	18.1	663	US-10-108-605-245	Sequence 245, App
17	598.5	18.0	674	US-10-090-455-4	Sequence 4, Appl1
18	590.5	17.8	638	US-10-072-621-10	Sequence 10, Appl1
19	585.5	17.6	646	US-10-072-621-9	Sequence 9, Appl1

20	585.5	17.6	646	US-10-090-455-2	Sequence 2, Appl1
21	578.5	17.4	627	US-10-090-455-8	Sequence 8, Appl1
22	578	17.4	604	US-09-745-763-197	Sequence 197, App
23	570.5	17.2	646	US-10-154-452-4	Sequence 4, Appl1
24	567.5	17.1	646	US-10-090-455-13	Sequence 13, Appl1
25	562.5	16.9	646	US-10-154-452-8	Sequence 8, Appl1
26	485	14.6	545	US-10-083-357-1335	Sequence 1335, App
27	403.5	12.1	1564	US-09-801-368-244	Sequence 244, App
28	398	12.0	1501	US-09-801-368-346	Sequence 346, App
29	374	11.2	1511	US-09-801-368-250	Sequence 250, App
30	270.5	8.1	306	US-09-738-626-4554	Sequence 4554, App
31	268	8.1	242	US-09-769-787-83	Sequence 83, Appl1
32	263.5	7.9	345	US-09-815-242-11068	Sequence 11068, A
33	263	7.9	241	US-09-815-242-11194	Sequence 11194, A
34	241.5	7.3	247	US-09-738-626-5108	Sequence 5108, App
35	240	7.2	1279	US-10-097-340-2	Sequence 2, Appl1
36	240	7.2	1280	US-10-072-621-7	Sequence 7, Appl1
37	240	7.2	1280	US-09-866-866A-2	Sequence 2, Appl1
38	240	7.2	1280	US-09-866-866A-4	Sequence 4, Appl1
39	239.5	7.2	2436	US-10-156-239-8	Sequence 8, Appl1
40	239.5	7.2	2436	US-10-199-485-8	Sequence 8, Appl1
41	239.5	7.2	2436	US-09-795-693-8	Sequence 8, Appl1
42	238.5	7.2	1594	US-09-971-121-4	Sequence 4, Appl1
43	238.5	7.2	1642	US-09-971-121-2	Sequence 2, Appl1
44	236.5	7.1	385	US-09-769-787-144	Sequence 144, App
45	236	7.1	203	US-10-252-819-17	Sequence 17, Appl1

ALIGNMENTS

RESULT 1
US-10-090-455-6
; Sequence 6, Application US/10090455
; Publication No. US20030027259A1
GENERAL INFORMATION:
APPLICANT: Chen, Hongyun
APPLICANT: Le Bihan, Stephane
TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
FILE REFERENCE: 100103 406
CURRENT APPLICATION NUMBER: US/10/090,455
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 651
TYPE: PRT
ORGANISM: Homo sapiens
US-10-090-455-6

Query Match 100.0%: Score 3326; DB 9; Length 651;
Best Local Similarity 100.0%: Pred. No. 1.6e-285;
Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MGDLSTLFGSGMGLOVNRGSSQSLGAPATAPEPHSLGILHASYSVSHRVPMDITSC	60
DB	1	MGDLSTLFGSGMGLOVNRGSSQSLGAPATAPEPHSLGILHASYSVSHRVPMDITSC	60
QY	61	ROQWROIILKDYSLVYESQVLCIIGSSGRTTLLDAMSGRAGTFLGEVYNGRAL	120
DB	61	ROQWROIILKDYSLVYESQVLCIIGSSGRTTLLDAMSGRAGTFLGEVYNGRAL	120
QY	121	REQOPDEFSYVLOSITLSSLTVEETLHYTALLAIRGNSPQKQVAVNAELSLSHV	180
DB	121	REQOPDEFSYVLOSITLSSLTVEETLHYTALLAIRGNSPQKQVAVNAELSLSHV	180
QY	181	ADRLIGNTSLGISTGERRRYSIAQLQDPVLMFDEPTTGLDQMTANOIVLLVELAR	240
DB	181	ADRLIGNTSLGISTGERRRYSIAQLQDPVLMFDEPTTGLDQMTANOIVLLVELAR	240
QY	241	RNRIVVLTIHOPRSELPQFLDKIALTLRGLIEGTPADMUDFNDCGYPCEHSPNPEDE	300
DB	241	RNRIVVLTIHOPRSELPQFLDKIALTLRGLIEGTPADMUDFNDCGYPCEHSPNPEDE	300

Query Match	100.0%	Score 3326	DB 9	Length 651
Best Local Similarity	100.0%	Pred. No. 1,66-289		
Matches	651	Conservative 0	Mismatches 0	Indels 0
			Gaps 0	

QY	1	MGDLSSLTTPGSGMKLQVNRGSSSLLEGAPAPAPEPHSGILHASTSVSHRVRPMDITSC	60
Db	1	MGDLSSLTTPGSGMKLQVNRGSSSLLEGAPAPAPEPHSGILHASTSVSHRVRPMDITSC	60
QY	61	ROOFTROI LKDVSLYVESGQIMCTILGSSGSKTTLDDMSGRILGRAGTFLEGVYNGRAL	120
Db	61	ROOFTROI LKDVSLYVESGQIMCTILGSSGSKTTLDDMSGRILGRAGTFLEGVYNGRAL	120
QY	121	RREDFOCFESVYLQSDPLTSLSYATRETHYFALLAIRGNPSPQKRYEAVVAEISLSHY	180
Db	121	RREDFOCFESVYLQSDPLTSLSYATRETHYFALLAIRGNPSPQKRYEAVVAEISLSHY	180
QY	181	ADRLIGNTSLAGISTGERRRYSIAQLIDPKVLFDEPTTGIDCMTANQIVLLVELAR	240

[illegible]

121 RRRQFDDCFSTYVLSQSTLSSLTFRVETLHTALLARKNGPSFOKKEAVNAEISLSH 180
DB
181 ADRLLGNYSLSGISTGERRRVSIAAQLDDPKVMEDEPTTGDCMTANOIVLVEIAR 240
OY
181 ADRLLGNYSLSGISTGERRRVSIAAQLDDPKVMEDEPTTGDCMTANOIVLVEIAR 240
DB
241 RRRIVVLTTHQPRSELFQFDKIALISFGBELIFCGTPAEMLDFNDCGYPCEHSNPDF 300
OY
241 RRRIVVLTTHQPRSELFQFDKIALISFGBELIFCGTPAEMLDFNDCGYPCEHSNPDF 300
DB
301 YMDLTSVDTQSKERELETSEKRVOMIESAKKSAICHTLKNIERMKHLTLPMPVFKTK 360
OY
301 YMDLTSVDTQSKERELETSEKRVOMIESAKKSAICHTLKNIERMKHLTLPMPVFKTK 360
DB
361 SPQVSKLGLVLRVTRNLVRNKLAVITRLQNLINGELFLFVLVRVSNVLKGAIQDRV 420
OY
361 SPQVSKLGLVLRVTRNLVRNKLAVITRLQNLINGELFLFVLVRVSNVLKGAIQDRV 420
DB
421 GLLYQVCGATPTTGMLNANVLPVLRVAVSDQSDGLYOKQOMLALAYLHVPESVATM 480
OY
421 GLLYQVCGATPTTGMLNANVLPVLRVAVSDQSDGLYOKQOMLALAYLHVPESVATM 480
DB
421 GLLYQVCGATPTTGMLNANVLPVLRVAVSDQSDGLYOKQOMLALAYLHVPESVATM 480
OY
481 IFSSVCYMTLGLHPEVARGYFSALLAHLIGELFLVLLGIYONPNIVNSVVALLSIA 540
DB
481 IFSSVCYMTLGLHPEVARGYFSALLAHLIGELFLVLLGIYONPNIVNSVVALLSIA 540
OY
541 GVLVSGFLRNIOEMPIPKIISYTFQKYCSEILVNEFYGILNFCGSSNSVVTNPMK 600
DB
541 GVLVSGFLRNIOEMPIPKIISYTFQKYCSEILVNEFYGILNFCGSSNSVVTNPMK 600
OY
601 AFTOGIOFIEKTCPCGATSRFTANFLILYSFIPALVILGIVFKIRDLISR 651
DB
601 AFTOGIOFIEKTCPCGATSRFTANFLILYSFIPALVILGIVFKIRDLISR 651

Result 4

US-09-989-981a-2

Sequence 2, Application US/09989981A
Publication No. US20030049730A1
GENERAL INFORMATION:
APPLICANT: Hobbs, Helen H.
APPLICANT: Shan, Bel
APPLICANT: Barnes, Robert
APPLICANT: Tian, Hui
APPLICANT: Tularik Inc.
APPLICANT: Board of Regents, The University of Texas System
TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
FILE REFERENCE: 018781-007320US
CURRENT APPLICATION NUMBER: US/09/989,981A
CURRENT FILING DATE: 2002-07-23
PRIOR APPLICATION NUMBER: US 60/252,235
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/253,645
PRIOR FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 652
TYPE: PRT
ORGANISM: Mus musculus
FEATURE:
OTHER INFORMATION: mouse ABCG5 (MABG5)
US-09-989-981a-2

Query Match 82.5%; Score 2744.5; DB 9; Length 652;
Best Local Similarity 80.2%; Pred. No. 2.2e-237;
Matches 523; Conservative 64; Mismatches 64; Indels 1; Gaps 1;
OY 1 MGDLSLTPTGGSGMGLQVNGSOSLSGAPATAPET-HSGLIHAHASYSHRVPMWDITS 59
DB 1 MGELFISPEGARGHINNGSLSSLEQSVTCTEARRHSLGVLAHVSYSNKRKGPMMNKS 60

60 CRQOFTROLKDVSLYVESGOIWCITLSSGSGCKTLLDAMSGFLGAGFLGEVYNGRA 119
DB
61 COOKMDROILKDVSLYIESGOIWCITLSSGSGCKTLLDAMSGFLGAGFLGEVYNGRA 120
OY
120 LRREPODCEFSYVLSQSTLSSLTFRVETLHTALLARKNGPSFOKKEAVNAEISLSH 179
DB
121 LRREPODCEFSYVLSQSTLSSLTFRVETLHTALLARKNGPSFOKKEAVNAEISLSH 180
OY
180 VADRLLGNYSLSGISTGERRRVSIAAQLDDPKVMEDEPTTGDCMTANOIVLVEIAR 239
DB
181 VADRLLGNYSLSGISTGERRRVSIAAQLDDPKVMEDEPTTGDCMTANOIVLVEIAR 240
OY
240 RRRIVVLTTHQPRSELFQFDKIALISFGBELIFCGTPAEMLDFNDCGYPCEHSNPDF 299
DB
241 RRRIVVLTTHQPRSELFQFDKIALISFGBELIFCGTPAEMLDFNDCGYPCEHSNPDF 300
OY
300 FYMDLTSVDTQSKERELETSEKRVOMIESAKKSAICHTLKNIERMKHLTLPMPVFKTK 359
DB
301 FYMDLTSVDTQSKERELETSEKRVOMIESAKKSAICHTLKNIERMKHLTLPMPVFKTK 360
OY
360 DSGVFSKLGVLRRVTRNLVRNKLAVITRLQNLINGELFLFVLVRVSNVLKGAIQDRV 419
DB
361 DSGVFSKLGVLRRVTRNLVRNKLAVITRLQNLINGELFLFVLVRVSNVLKGAIQDRV 420
OY
420 VGLLYQVCGATPTTGMLNANVLPVLRVAVSDQSDGLYOKQOMLALAYLHVPESVAT 479
DB
421 VGLLYQVCGATPTTGMLNANVLPVLRVAVSDQSDGLYOKQOMLALAYLHVPESVAT 480
OY
480 MIFSSVCYMTLGLHPEVARGYFSALLAHLIGELFLVLLGIYONPNIVNSVVALLSI 539
DB
481 MIFSSVCYMTLGLHPEVARGYFSALLAHLIGELFLVLLGIYONPNIVNSVVALLSI 540
OY
540 AGVLVSGFLRNIOEMPIPKIISYTFQKYCSEILVNEFYGILNFCGSSNSVVTNPMK 599
DB
541 SGLLISGFLRNIOEMPIPKIISYTFQKYCSEILVNEFYGILNFCGSSNSVVTNPMK 600
OY
600 CAFTOGIOFIEKTCPCGATSRFTANFLILYSFIPALVILGIVFKIRDLISR 651
DB
601 CAFTOGIOFIEKTCPCGATSRFTANFLILYSFIPALVILGIVFKIRDLISR 652

Result 5

US-09-837-992-1

Sequence 1, Application US/09837992
Patent No. US20020081687A1
GENERAL INFORMATION:
APPLICANT: Tian, Hui
APPLICANT: Schultz, Joshua
APPLICANT: Shan, Bel
APPLICANT: Tularik Inc.
TITLE OF INVENTION: Stossterolemia Susceptibility Gene (SSG): Compositions
and Methods of Use
FILE REFERENCE: 018781-006020US
CURRENT APPLICATION NUMBER: US/09/837,992
CURRENT FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: US 60/198,465
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: US 60/204,234
PRIOR FILING DATE: 2000-05-15
NUMBER OF SEQ ID NOS: 45
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 652
TYPE: PRT
ORGANISM: Mus musculus
FEATURE:
OTHER INFORMATION: mouse stossterolemia susceptibility gene (SSG)
US-09-837-992-1

Query Match 82.5%; Score 2744.5; DB 10; Length 652;
Best Local Similarity 80.2%; Pred. No. 2.2e-237;
Matches 523; Conservative 64; Mismatches 64; Indels 1; Gaps 1;

RESULT 6
US-09-989-981A-4
; Sequence 4, Application US/09989981A
; Publication No. US20030049730A1

```

1  APPLICANT:  Tidalix Inc.
2  APPLICANT:  Board of Regents, The University of Texas System
3  TITLE OF INVENTION:  ABCG5 and ABCG8:  Compositions and Methods of Use
4  FILE REFERENCE:  018781-007320US
5  CURRENT APPLICATION NUMBER:  US/09/989,981A
6  CURRENT FILING DATE:  2002-07-23
7  PRIOR APPLICATION NUMBER:  US 60/252,235
8  PRIOR FILING DATE:  2000-11-20
9  PRIOR APPLICATION NUMBER:  US 60/253,645
10 PRIOR FILING DATE:  2000-11-28
11 NUMBER OF SEQ ID NOS:  13
12 SOFTWARE:  PatentIn Ver. 2.1
13 SEQ ID NO 4
14 LENGTH:  672
15 TYPE:  PRT
16 ORGANISM:  Mus musculus
17 FEATURE:
18 OTHER INFORMATION:  mouse ABCG8 (mABCG8)

```

```

0Y 15 LONRGSSSLLECAPYTAPEPHSLGILHAYSVSJHRV-----PWMD-ITSCR 61
Db 17 LODASSGLQDLSL-----PSSSEDNLSVFTYSGGQNTLEVRDLTYOVDIASOYWHFEDLQFK 72
0Y 62 QOWTRQI-----LKDVSLVYESQIMCLIGSSGSGKTTLLDMSGRILRACFT-IGE 113
Db 73 IPMHSSSDSCELGIRLNLSPKRSOGOMALIIIGSSCCGRASLLDVTYTR-CHGGKMSGQ 131
0Y 113 VYVGRALRREOFODEFSYVLOSDTLLSLTYRETLHTYALAI-IRBNPSPFKVRYAV 177
Db 132 IWMNGOSTQLYRKCYAHYRQHODLLPMLTYRETLTAQMRLEPRTTSQARODRKEDEV 199
0Y 172 MAELSLSHVADRLIGNTSIGISITGBRRRVSIAAOLDDPKVMLFDEPTTGIDCMTANOI 231
Db 192 IAEIRLRQCANTHRVGTNYRGVSGGERRRVSIGYOLLMPGILLDEPTSGDSPYAHNL 25
0Y 232 VVLLLEARNRIRVLTITQPRSELFOQLDAISFGBELFCGTPRABMLDFPNDGCTPC 299
Db 252 VTTLSIRLAKCNRVLYLSLHQPRSDIRFRLDVLMTSGTPYLGAAQOMVOYFTFSIHP 311
0Y 292 PHSNPDEPFMDLTSVDQSKEREIETSKRYOMIESAKKSA-----ICAKTLKNIERM 345
Db 312 PRSNNPADEFVDLTSLDRSKREBEVATYKQASIALALEKAYOGDDPLTMAEKEINTS 371
0Y 346 KHLKTLPMVDFKTKDS-----PGVFSKIGVLLRVRTRLVRNKLAVITRLLQNTIMG 397
Db 372 THYVLSLTL-----TQDDCGSTAVELPQMIQFSTLIRROISMDFRDPLTLIHGSEACIMS 427
0Y 398 LFTLFFLVRYSVNLGALIDRGLLYOYGVAFTPTGLMLNVNLFVYLRAVSDQSDGL 457
Db 428 LIIGFLYLYGGNAQL--SEMDRALLLFMI GALIRPNVILVDVYSKCHSRKSRMYTELEGL 485
0Y 458 YOKROMMLAVLALHVPFSVVAATMIFSSVCYWTGLHPEVARFGYFSALLAPHLIGFL- 516
Db 486 YTAGPPEFFAKILGELBEHCAYVILIYAMPYIWTNLRPPELF-----LT--HFLVWLV 537
0Y 517 -----TLYVLGIYQONNI--VNSVVALLSIAGLVSSGGLRNIOEMPIKIIISFTQKY 570
Db 538 VECRCRIMALAASAMLPTEFHMSSEFCFCALYNSFYTLAGMINDIMLMIYPAWISKLSLRLW 597
0Y 571 CSBILVYNEFYGLNFT--CGSSNVSVTTPMPCAFYQIOFTEKTCPGATSRFTNMFILY 628
Db 598 CFSGLMQIOENGLHYLTQIGNFTPSLDDTM-----ISAMDNSHPLY 640
0Y 629 SFTPALVYIIGI 639
Db 641 AIT--LIVIGI 649

```

```

RESULT 7
US-10-090-455-7
: Sequence 7, Application US/10090455
: Publication No. US20030027259A1
: GENERAL INFORMATION:
: APPLICANT: Chen, Hongyun
: APPLICANT: Le Bihan, Stephanie
: TITLE OF INVENTION: NOVEL ABC4 TRANSPORTER AND USES THEREOF
: FILE REFERENCE: 100103.406
: CURRENT APPLICATION NUMBER: US/10/090,455
: CURRENT FILING DATE: 2002-03-01
: NUMBER OF SEQ. ID NOS: 17
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ. ID NO. 7
: LENGTH: 673
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-10-090-455-7

```

Query Match	21.0%;	Score 697;	DB 9;	Length 673;
Best Local Similarity	28.9%;	Pred. No. 9.5e-54;		
Matches 187;	Conservative 124;	Mismatches 241;	Indels 96;	Gaps 16;

```

0Y 8TFGSGSGLOAVNRNOSGSLBGAPAT-APEHSHGILHAASYSHRVR-PMWD-ITSCQOW 64
   ||:||||
16 TPODTSGLODRLEFSSSDSNLSYFTTYSGQPTLEVRDLNTOVNDIASOVPMBEOLQOFMPW 75
   ||:||||
0Y TROI-----LKDVSLYESGQIMCILGSSGSGKTTLLDAMSGRLGRGTF-IGEVY 115
   |
Db TSPSCONSCSELIGQNLISFKYVRSGOMLAIIGSSCCGRASILDYTR- GHGCKIKSQIM 134
   ||:||||
0Y 116NGRALRREGFOCFVYLOSDTLSSLYRETIHATALLAI-RKNGBSQOKVEAVMAE 174
   ||:||||
Db 135NQPSSPOLVRCVAVHROHNDLPLVYRETIATLAOMKLPTFQOARDKREVEDIAE 194
   ||:||||
0Y 175LSLSHVADBLIGNYSIGSIGTGERRRVSTAIOLODPKWLFGPEPTGDCMANTIVYL 234
   ||:||||
Db 195LRLRQCADPRVGNMRYRGISGGERRRVSTGYOLNMPGILIIDEPTSGDLSFAHMLVKT 254
   ||:||||
0Y 235LYELARRNRIVVLTTHQPSLELPQLEFDKTAIISFGELIFCGPRAEMIDFENDGYPDEH 294
   ||:||||
Db 255LSRLAKGNLTVLISLHOPRSDIFRFLDVLVLTGSTPIYLGAAOHVQYFTAJPCPRY 314
   ||:||||
0Y 295SNPFYVDLTGVDPNOSKREIETSKRYOMITSAYKKSAICHTLNIEMKHL----- 348
   ||:||||
Db 315SNRADFYDLTSDIRRSRQEOLATREKAOISLALF-----LEVRDLDFLWK 362
   ||:||||
0Y 369-----FTLPM-----VPEKTKDSGVSFKLGVLLRRTYRNLVNRMLAVYTRL 390
   ||:||||
Db 363AETKDLDEDTCVSSSVTPLDTCNCLPSPTK-MGAVQOFTTLRLRQISNDRPDLPTLIHG 421
   ||:||||
0Y 391LQNLINGLELFEVLARVSNVLKGAIQ----DRVGLLYOFVGATPYTGMLNAYNLFPVLR 446
   ||:||||
Db 422AEACIMSMTIGLRYG-----HGSIOFSMDTALLFMGALIPRNVLIDVSCYCSER 475
   ||:||||
0Y 447AVSDQESDGLQOKQOMLAVLAHLVPESVATMTFSSVCYMTGLGHPVAR----- 499
   ||:||||
Db 476AMLYELRDLGIXTTPPYFAKLTIGELPEICAIYIIIGMPTMLANLRPGIOPLFLHFLV 535
   ||:||||
0Y 500-----GTFSAALLAPHLIGELFLVLLGIQVONPNIVNSVALLSIAGVLVSGFL 549
   ||:||||
Db 536WLVPVFCRIMALAALLPTFHMAFFS-----NALYNSFYLAG-----GFM 577
   ||:||||
0Y 550RNIDEMPIPKIISYTFQOKCESELIVNVEFGIAMPCTGSSSVSYTIN 597
   ||:||||
Db 578INLSLMKTVPAKISVSEFLRWCFEGLMKIQESRRRYKMDPLGNITAAVS 625

```

```

RESULT 8
US-09-989-981A-8
: Sequence 8, Application US/09989981A
: Publication No. US20030049730A1
: GENERAL INFORMATION:
: APPLICANT: Hobbs, Helen H.
: APPLICANT: Shan, Bel
: APPLICANT: Barnes, Robert
: APPLICANT: Tian, Hui
: APPLICANT: Tularik Inc.
: TITLE OF INVENTION: Board of Regents, The University of Texas System
: FILE REFERENCE: 018/781-00732005
: CURRENT APPLICATION NUMBER: US/09/989,981A
: CURRENT FILING DATE: 2002-07-23
: PRIOR APPLICATION NUMBER: US 60/252,235
: PRIOR FILING DATE: 2000-11-20
: PRIOR APPLICATION NUMBER: US 60/253,645
: PRIOR FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 13
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 8
: LENGTH: 673

```

```

: TYPE: prt
:
: ORGANISM: Homo sapiens
:
: FEATURE:
:
: OTHER INFORMATION: human ABCG8 (hABCG8)
US-09-989-981A-8

```

Query Match	21.08;	Score 697;	DB 9;	Length 673;
Best Local Similarity	28.94;	Pred. No. 9.5e-54;		
Matches 187;	Conservative 124;	Mismatches 241;	Indels 96;	Gaps 16;

```

QY      8 TPGSGMGLDVNGSSGSSLEGAPAT-APRPHSIGILHMASYSHRVR-PMWD-TSCNQW 64
Db      16 TPQDTSGLDRLFSSSESNSLFTYTSQGPNTLEVRDLNTQYDLASQVPHFEQLAQEFKPM 75
QY      65 TROI-----LKDVSLYVESGQIMCITLSSGSGKTLTLDAMSGRYGRGTF-IGEVY 115
Db      76 TSPSCONSCGLDIONLSEFKVRSGOMLAIITGSSCGRASLDVITGR-GHGKIKSGQIM 134
QY      116 NGCALREPOFCFSYVLQSDTLSSLTVRRTLHTHTALTA-RGNGCSGQKKEAVMAE 174
Db      135 NGQSSPOLYRKCVAHVQHQNQLPNTLVRETLFAIQMRLPFEISOQRKREVDIAE 194
QY      175 LSLSHVADLIYNISLIGISTGERRRVASIAQLODDPKMLDEPTTGLCOMANTQIVL 234
Db      195 LRLRQCADPRVGNMVRIGLSGGERRRVASIQVQLMNPGLILDLDEPTSGDLSFRANLVKT 254
QY      235 LVELARRNIYVLTTHQPSBELFQLEDKIALISFQELFCGTPAEMLDFPNDGYPCEH 294
Db      255 LSLRLKGNLTVLISLHQRSDIFRFLDLYLLMTSGTPYTLGAOHMAYFTALGPCRY 314
QY      295 SNPFEDYMLDVTVDOSKERETESKRWOMESAYKKSAIOHKLKLNIEPMKH- 348
Db      315 SNRPADYVDTLSIDRRSKRELATREKQSLAUF-----LEVRDLDLFLMK 362
QY      349 -----KTLPM-----VPEFKRDSPPYFSKGLVLRREVTENLVNKLAVYTRL 390
Db      363 AETKDLDEDDTVESSVTPFLDNTCLPSPTK-MPGAVQGFPTLLIRQISNDPRDLPTLLIHG 421
QY      391 LQNLIMGFLFLEFVLVRASNVLKALIQ-----DRGGLLYQYVGATPYTGMTANAVNLPVLR 446
Db      422 ABRCLASMTWIGFLYFG-----HQSIOISFMDTALLFMIGALIPFVNIIDVISCYSER 475
QY      447 AVSDSESDQGLQOKOMMLAYALAHVLPFSVAATIFSSVCWTTGLGLEPVARF----- 499
Db      476 AMLYLELEBGLYTTQPPFAKILGELPBHCAYIIITIGMPYTWLANLRPGLOPPELLHLLY 535
QY      500 -----GYSAALLAPHLIGELFTLVLLGIVQNPNTVNSVALLSIAGVLVYSGFL 549
Db      536 WLVEFCRIMALAAALLPTFHMASFS-----NALYNSFYLAG-----GFM 577
QY      550 RNIGQEMPIPKIISYFTQKCSCELYVNEGYTGFPGGSSNNVSVTYN 597
Db      578 ILLSSITVVPAMISVSLKWCFCGLMKIQSSRRTPYKPLGNTLIAVS 625

```

RESULT 9
 US-09-961-086-1
 ; Sequence 1, Application US/09961086
 ; Publication No. US20030036645A1
 ;
 GENERAL INFORMATION:
 ; APPLICANT: UNIVERSITY OF MARYLAND, BALTIMORE
 ; APPLICANT: ROSS, Douglas D.
 ; APPLICANT: DOYLE, L. Austin
 ; APPLICANT: ABRUZZO, Lynne
 ; TITLE OF INVENTION: BREAST CANCER RESISTANCE PROTEIN (BCRP) AND THE DNA
 ; TITLE OF INVENTION: WHICH ENCODES IT
 ; FILE REFERENCE: EP19376-019
 ; CURRENT APPLICATION NUMBER: US/09/961,086
 ; CURRENT FILING DATE: 2001-09-21
 ; PRIOR APPLICATION NUMBER: US 60/073,763
 ; PRIOR FILING DATE: 1998-02-05
 ; PRIOR APPLICATION NUMBER: PCT/US99/02577
 ; PRIOR FILING DATE: 1999-02-05

; PRIOR FILING DATE: 2000-10-06
 ; PRIOR APPLICATION NUMBER: US 09/731261
 ; PRIOR FILING DATE: 2000-12-06
 ; NUMBER OF SEQ ID NOS: 61
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 61
 ; LENGTH: 655
 ; TYPE: PRF
 ; ORGANISM: Homo sapiens
 US-10-120-687-61

Query Match 20.5%; Score 680.5; DB 9; Length 655;
 Best Local Similarity 29.2%; Pred. No. 2,7e-52;
 Matches 182; Conservative 137; Mismatches 250; Indels 55; Gaps 18;

QY 21 SOSLEGAPATAP---EPHSLGILHASYSVSHRVRPMMDITSCROQMTROLTKDVSLEYE 77
 Db 13 SGGTNGFPATVSDKLAFTGAVLSFHNTICRYKLKSGFLPCRKPEKEILSNINGIMK 72
 QY 78 SGOIMCIISSGSGKTTLLDAMSGRLGACFTFLGEEVYVNGRALRREOFODCFSTVYLOSDT 137
 Db 73 PG-LNALIGPTGGKSSLDLVLAARKDPSG-LSGDVLING-APRPANFKCSGYVODDV 129
 QY 138 LLSLTYRETLHATALLAIRGNPG-SFOKKVEAVMAELSLSHVADRILGNSLGISTG 196
 Db 130 VMGTLTYREMLQFSAALRLATYTNHNEKNERINRVIGELGLDKYADSKVGTQFTRGYSGG 189
 QY 197 ERRRVSTAAOLLQDPKWLDEPTTGLDCMTANOIVLVLELARNRHIVYLTTHQPSSEL 256
 Db 190 ERKRTSIGMELITDPSILFDEPTTGLDSTANAVLILLRMSKQGRITIFSHQPRYSI 249
 QY 257 FOLFDKTALLSPGELICGTPAEMLDPEFNDGYPCEPHSNPFDEYMDLTSVDTO----SK 312
 Db 250 FKLFDLSLTLLASGRMLFHPAQAELGYFESAGYCEAYNNPADFDLIIINGDSTAVALNR 309
 QY 313 ERE-----IETSKR-----VOMIESAVKSAICHKT-----LKNIERMKHLKTLPAVPF 356
 Db 310 EEDFKATEIIEPSKODPLIEKLAETIVNSFYETKAEHLQOLSGGKKKIYFKEIYS 369
 QY 357 KTKDSPGVESKGLVLRVTRNRLVNRNLAVITRLLQNLINGLFL--LEFVLVRANSYVLKG 414
 Db 370 TT-----SFCHQLRNVSRSRKKNLGNQASIAQIIVTVGLVIGALYFGKNDST---- 421
 QY 415 AIDRVGLLYQFVATPYTGMLNANVLPVLAVSODESDGLYQKQMMALAYAL-HVLP 473
 Db 422 GIONRAGVLP-FLITNOCFSSVSVELFVEYKELFHEIYISGYRVSSEYFLGKLSDLLP 480
 QY 474 FSVVATMIFSSVCYTWLGLHPEVARFGYFSALLAPHLIGBFLTVLLGIVQNEINYSV 533
 Db 481 MRMLPSIIFTCIYFMLGLKPKADAFVMMFTL---MVAYSASSMALALAIAGOSVVA 537
 QY 534 VALISTAGV--LVGSGFLRNIOEMPIPKIISYFTPOKCYSEILVNEFGLNFTCCSSN 591
 Db 538 TLMATICFVMMIFSGLLVNLTTIASWLSMLOYFSIPRYGFTALQHNHPELGQNCPC--- 594
 QY 592 VSVTTNPMCAFTGIGIETKTCPG 615
 Db 595 LNAITGNMPCMYA-----TCYG 610

RESULT 12
 US-10-090-455-5
 ; Sequence 5, Application US/10090455
 ; Publication No. US20030027259A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Chen, Hongyun
 ; APPLICANT: Le Bihan, Stephane
 ; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
 ; FILE REFERENCE: 100103.406
 ; CURRENT APPLICATION NUMBER: US/10/090.455
 ; PRIOR FILING DATE: 2002-03-01
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5
 ; LENGTH: 655
 ; TYPE: PRF
 ; ORGANISM: Homo sapiens
 US-10-090-455-5

Query Match 20.3%; Score 674.5; DB 9; Length 655;
 Best Local Similarity 29.0%; Pred. No. 9.5e-52;
 Matches 181; Conservative 137; Mismatches 251; Indels 55; Gaps 18;

QY 21 SOSLEGAPATAP---EPHSLGILHASYSVSHRVRPMMDITSCROQMTROLTKDVSLEYE 77
 Db 13 SGGTNGFPATVSDKLAFTGAVLSFHNTICRYKLKSGFLPCRKPEKEILSNINGIMK 72
 QY 78 SGOIMCIISSGSGKTTLLDAMSGRLGACFTFLGEEVYVNGRALRREOFODCFSTVYLOSDT 137
 Db 73 PG-LNALIGPTGGKSSLDLVLAARKDPSG-LSGDVLING-APRPANFKCSGYVODDV 129
 QY 138 LLSLTYRETLHATALLAIRGNPG-SFOKKVEAVMAELSLSHVADRILGNSLGISTG 196
 Db 130 VMGTLTYREMLQFSAALRLATYTNHNEKNERINRVIGELGLDKYADSKVGTQFTRGYSGG 189
 QY 197 ERRRVSTAAOLLQDPKWLDEPTTGLDCMTANOIVLVLELARNRHIVYLTTHQPSSEL 256
 Db 190 ERKRTSIGMELITDPSILFDEPTTGLDSTANAVLILLRMSKQGRITIFSHQPRYSI 249
 QY 257 FOLFDKTALLSPGELICGTPAEMLDPEFNDGYPCEPHSNPFDEYMDLTSVDTO----SK 312
 Db 250 FKLFDLSLTLLASGRMLFHPAQAELGYFESAGYCEAYNNPADFDLIIINGDSTAVALNR 309
 QY 313 ERE-----IETSKR-----VOMIESAVKSAICHKT-----LKNIERMKHLKTLPAVPF 356
 Db 310 EEDFKATEIIEPSKODPLIEKLAETIVNSFYETKAEHLQOLSGGKKKIYFKEIYS 369
 QY 357 KTKDSPGVESKGLVLRVTRNRLVNRNLAVITRLLQNLINGLFL--LEFVLVRANSYVLKG 414
 Db 370 TT-----SFCHQLRNVSRSRKKNLGNQASIAQIIVTVGLVIGALYFGKNDST---- 421
 QY 415 AIDRVGLLYQFVATPYTGMLNANVLPVLAVSODESDGLYQKQMMALAYAL-HVLP 473
 Db 422 GIONRAGVLP-FLITNOCFSSVSVELFVEYKELFHEIYISGYRVSSEYFLGKLSDLLP 480
 QY 474 FSVVATMIFSSVCYTWLGLHPEVARFGYFSALLAPHLIGBFLTVLLGIVQNEINYSV 533
 Db 481 MRMLPSIIFTCIYFMLGLKPKADAFVMMFTL---MVAYSASSMALALAIAGOSVVA 537
 QY 534 VALISTAGV--LVGSGFLRNIOEMPIPKIISYFTPOKCYSEILVNEFGLNFTCCSSN 591
 Db 538 TLMATICFVMMIFSGLLVNLTTIASWLSMLOYFSIPRYGFTALQHNHPELGQNCPC--- 594
 QY 592 VSVTTNPMCAFTGIGIETKTCPG 615
 Db 595 LNAITGNMPCMYA-----TCYG 610

RESULT 13
 US-09-866-866A-10
 ; Sequence 10, Application US/09866866A
 ; Patent No. US20020102244A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sorrentino, Brian
 ; APPLICANT: Schuetz, John
 ; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
 ; FILE REFERENCE: 1340-1-021CIP2
 ; CURRENT APPLICATION NUMBER: US/09/866.866A
 ; PRIOR FILING DATE: 2001-08-30
 ; PRIOR APPLICATION NUMBER: 09/584,586
 ; PRIOR FILING DATE: 2000-05-31
 ; PRIOR APPLICATION NUMBER: PCT/US99/11825
 ; PRIOR FILING DATE: 1999-05-27
 ; PRIOR APPLICATION NUMBER: 60/086,988
 ; PRIOR FILING DATE: 1998-05-28
 ; NUMBER OF SEQ ID NOS: 27

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: June 11, 2003, 09:04:17 ; Search time 41 Seconds
(without alignments)
2115.760 Million cell updates/sec

Title: US-09-989-981A-6
Perfect score: 3326
Sequence: 1 MGDLSLTGGSMGLQVNRG.....PALVILGIYVEKIRDLISR 651

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
A.Genseq.101002:*

- 1: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1980.DAT.*
- 2: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1981.DAT.*
- 3: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1982.DAT.*
- 4: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1983.DAT.*
- 5: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1984.DAT.*
- 6: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1985.DAT.*
- 7: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1986.DAT.*
- 8: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1987.DAT.*
- 9: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1988.DAT.*
- 10: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1989.DAT.*
- 11: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1990.DAT.*
- 12: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1991.DAT.*
- 13: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1992.DAT.*
- 14: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1993.DAT.*
- 15: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1994.DAT.*
- 16: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1995.DAT.*
- 17: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1996.DAT.*
- 18: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1997.DAT.*
- 19: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1998.DAT.*
- 20: /SIDS2/gcgdata/genseq/genseqp-emb1/AA2000.DAT.*
- 21: /SIDS2/gcgdata/genseq/genseqp-emb1/AA2000.DAT.*
- 22: /SIDS2/gcgdata/genseq/genseqp-emb1/AA2001.DAT.*
- 23: /SIDS2/gcgdata/genseq/genseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3326	100.0	651	23	AAU96984	Human ABCG5 prote
2	3326	100.0	651	23	AAE13390	Human sitosterola
3	3323	99.9	651	23	AAU96989	Human ABCG5 mutat
4	3321	99.8	651	23	AAU96989	Human ABCG5 mutat
5	3321	99.8	651	23	AAU96989	Human ABCG5 mutat
6	3319	99.8	651	23	AAU96993	Human ABCG5 mutat
7	2748.5	82.5	652	23	AAE13309	Human ABCG5 mutat
8	2744.5	82.5	652	23	AAE13289	Mouse sitosterola
9	2742.5	82.5	652	23	AAE13308	Mouse sitosterola
10	2738.5	82.3	652	23	AAU96985	Mouse ABCG5 protei

XX	AC	AAU96984;
XX	DT	30-JUL-2002 (first entry)
XX	DE	Human ABCG5 protein.
XX	KW	Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol; arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease; chromosome 2p21.
XX	OS	Homo sapiens.
XX	FH	Key
FT	Misc-difference	Location/Qualifiers 2..15
FT	/note-	"Encoded by GGCCTC"
XX	PD	WO200227016-A2.
XX	PD	04-APR-2002.
XX	PF	25-SEP-2001; 2001MO-US29859.
XX	PR	25-SEP-2000; 2000US-235268P.
XX	PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.
XX	PA	(PATE/) PATEL S B.
XX	PA	(DEAN/) DEAN M.
PI		Patel SB, Dean M;

XX WPI: 2002-416483/44.
 DR N-PSDB: ABR51681.
 XX
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT strossterolemia, arteriosclerosis and heart diseases
 PS Claim 52; Page 35-36; 66pp; English.
 XX
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing strossterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 protein of the invention. This
 CC sequence is encoded by the human ABCG5 gene located on chromosome 2p21.
 XX
 SQ Sequence 651 AA:
 Query Match 100.0%; Score 3326; DB 23; Length 651;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGDISSITPGSGMGLQVNRSGSSLEGAPATAPAPPHSLGILHASYSVSHRVPWMDITSC 60
 DB 1 MGDISSITPGSGMGLQVNRSGSSLEGAPATAPAPPHSLGILHASYSVSHRVPWMDITSC 60
 QY 61 RQOKTROIILNDVSLYVSSGOIMCTISGSSGKTTLLDAMSGRGRAGTFTGEYVNRAL 120
 DB 61 RQOKTROIILNDVSLYVSSGOIMCTISGSSGKTTLLDAMSGRGRAGTFTGEYVNRAL 120
 QY 121 RREQFOQCFSVYLOSPTLSSLSYRETLHTTALAIRGNPGSFQKVEYVMAELSHV 180
 DB 121 RREQFOQCFSVYLOSPTLSSLSYRETLHTTALAIRGNPGSFQKVEYVMAELSHV 180
 QY 181 ADRLIGVSLAGISTGERRRVSIQAQLLQDPKVMLEDEPTTGDCMTANOIVLLELAR 240
 DB 181 ADRLIGVSLAGISTGERRRVSIQAQLLQDPKVMLEDEPTTGDCMTANOIVLLELAR 240
 QY 241 RNRIVVLTIHOPRSELFOLEDKTAIISFGLIFCGTAPKEMLDFFNDCGCPRESNPDF 300
 DB 241 RNRIVVLTIHOPRSELFOLEDKTAIISFGLIFCGTAPKEMLDFFNDCGCPRESNPDF 300
 QY 301 YMDLTVDYOSKEREIETSKRVOMIESAYKKSACHTLTNIEEMKILKTLPMVPEKTD 360
 DB 301 YMDLTVDYOSKEREIETSKRVOMIESAYKKSACHTLTNIEEMKILKTLPMVPEKTD 360
 QY 361 SPGVFSLGVLARRVTRNLVNRKLAIVTTRLLQNLIMGLFLFLFVLRVSNVLGAIDRV 420
 DB 361 SPGVFSLGVLARRVTRNLVNRKLAIVTTRLLQNLIMGLFLFLFVLRVSNVLGAIDRV 420
 QY 421 GLIYOFGATPYTGMNAVNLFPVLRVSDOESODGLYQKWMMLAALAVLPEFSVATM 480
 DB 421 GLIYOFGATPYTGMNAVNLFPVLRVSDOESODGLYQKWMMLAALAVLPEFSVATM 480
 QY 481 IFSSVCYMTLGLHEVARFGYFSNALLAPHLIGFELTVLGLYQNPNIYNSVVALLSIA 540
 DB 481 IFSSVCYMTLGLHEVARFGYFSNALLAPHLIGFELTVLGLYQNPNIYNSVVALLSIA 540

QY 541 GVLVSGFLLNIDEMPIPFKIIISFTFQKCYSEILVNEFYGLNFTCGSSNVSVTNPMC 600
 DB 541 GVLVSGFLLNIDEMPIPFKIIISFTFQKCYSEILVNEFYGLNFTCGSSNVSVTNPMC 600
 QY 601 AFTQGIQFIKTCGATSRFTMNFLLIXSTIPALYIIGIYVFKIRHLISR 651
 DB 601 AFTQGIQFIKTCGATSRFTMNFLLIXSTIPALYIIGIYVFKIRHLISR 651
 RESULT 2
 ID AAE13290 standard; Protein; 651 AA.
 AC AAE13290;
 DT 12-FEB-2002 (first entry)
 DE Human strossterolaemia susceptibility gene (SSG) protein.
 KW Human; strossterolaemia susceptibility gene; SSG; atherosclerosis;
 KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; therapy;
 KW gall stone; coronary heart disease; cardiovascular disease; arthritis;
 KW xanthoma; haemolytic anaemia; transgenic animal; chromosome 2p21.
 OS Homo sapiens.
 PN WC200179272-A2.
 PD 25-OCT-2001.
 PE 18-APR-2001; 2001WO-0512758.
 PR 18-APR-2000; 2000US-198465P.
 PR 15-MAY-2000; 2000US-204234P.
 PA (TULA-) TULARIK INC.
 PI Tian H, Schultz J, Shan B;
 DR WPI: 2002-017598/02.
 DR N-PSDB: AAD22009.
 PT Novel strossterolemia susceptibility gene polypeptide and
 PT polynucleotide, useful for screening a compound that increases the
 PT level of expression or activity of SSG polypeptide for treating
 PT sterol-related disorder
 PS Claim 19; Fig 8; 105pp; English.
 XX
 CC The invention relates to an isolated strossterolaemia Susceptibility Gene
 CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
 CC binding cassette (ABC) family cholesterol transporter. SSG is useful
 CC for identifying a compound useful in the treatment or prevention of a
 CC sterol-related disorder, including strossterolaemia, hyperlipidaemia,
 CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or
 CC nutritional deficiencies. SSG is also useful for treating cholesterol-
 CC associated diseases or conditions including coronary heart disease and
 CC other cardiovascular diseases, and strossterolaemia-associated condition
 CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
 CC expression cassette is useful in the production of transgenic non-human
 CC animals. SSG genes and their homologues are useful as tools for a number
 CC of applications including diagnosing strossterolaemia and other
 CC cardiovascular disorders, for forensics and paternity determinations,
 CC and for treating any of a large number of SSG associated diseases. The
 CC present sequence is human SSG protein. Human SSG is located on chromosome
 CC 2p21.
 XX
 SQ Sequence 651 AA:
 Query Match 100.0%; Score 3326; DB 23; Length 651;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MGLSSLTGPGSMGLQVNRGSSQLSGCAPATAPRPHSIGILHASYSVSHRVRPMMDITSC 60
Db 1 MGLSSLTGPGSMGLQVNRGSSQLSGCAPATAPRPHSIGILHASYSVSHRVRPMMDITSC 60
QY 61 RQOMTROIILKDVSLVYESGQIMCIIIGSSGSKTTLDDAMSGRLGAGTFLGEVYVNGRAL 120
Db 61 RQOMTROIILKDVSLVYESGQIMCIIIGSSGSKTTLDDAMSGRLGAGTFLGEVYVNGRAL 120
QY 121 RRQFODCFSSVYVQSDTLSSLTAVRETHTYTTALAIRGNPGSFQKVEAVMAELSLSHV 180
Db 121 RRQFODCFSSVYVQSDTLSSLTAVRETHTYTTALAIRGNPGSFQKVEAVMAELSLSHV 180
QY 181 ADRLIGNYSLGISTGERRRVSTIAAQLDPPKVMLEPDEPTGIDCMANOIVLVLELAR 240
Db 181 ADRLIGNYSLGISTGERRRVSTIAAQLDPPKVMLEPDEPTGIDCMANOIVLVLELAR 240
QY 241 RNRIVVLTIHQPSSELFQFLDKITAILSFGLIFCGTFAEMLDFFNDCGYPCEHSNPDEF 300
Db 241 RNRIVVLTIHQPSSELFQFLDKITAILSFGLIFCGTFAEMLDFFNDCGYPCEHSNPDEF 300
QY 301 YMDLTSVDIOSKEREIETSKRVOMIESAVKSAICHKTLKNIEMKHLKTLPMVPFKTKD 360
Db 301 YMDLTSVDIOSKEREIETSKRVOMIESAVKSAICHKTLKNIEMKHLKTLPMVPFKTKD 360
QY 361 SPGVFSKLGVLRRVTRNLVNRKLAIVTTRLLQNLIMGLFELFVLRVRSNVLKGAIDRV 420
Db 361 SPGVFSKLGVLRRVTRNLVNRKLAIVTTRLLQNLIMGLFELFVLRVRSNVLKGAIDRV 420
QY 421 GLIXQVGAATPYTGMLNANVLFVLRVAVSQESQDGLYQKWMQMLAVLHLEPSSVATM 480
Db 421 GLIXQVGAATPYTGMLNANVLFVLRVAVSQESQDGLYQKWMQMLAVLHLEPSSVATM 480
QY 481 IFSSVCYMTLGLHPEVARFGYFSALLAPHLIGEFITLVILGIVONPNIVNSVALLSIA 540
Db 481 IFSSVCYMTLGLHPEVARFGYFSALLAPHLIGEFITLVILGIVONPNIVNSVALLSIA 540
QY 541 GVLVSGFLRNIOEMDIPKRIISYFTFOKYSEILVNEFYGLNFTCGSSNVSVTNNPMC 600
Db 541 GVLVSGFLRNIOEMDIPKRIISYFTFOKYSEILVNEFYGLNFTCGSSNVSVTNNPMC 600
QY 601 AFTOGIOFTEKTCRGATSRFTMNFLLISFTIPALVILGIVKIRIDHLISR 651
Db 601 AFTOGIOFTEKTCRGATSRFTMNFLLISFTIPALVILGIVKIRIDHLISR 651

RESULT 3
AA096992
ID AA096992 standard; Protein; 651 AA.
AC AA096992;
XX
XX
DT 30-JUL-2002 (first entry)
XX
XX
DE Human ABCG5 mutant E146Q protein sequence.
XX
XX
KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
XX mutant; muteln.
XX
OS Homo sapiens.
OS
OS Synthetic.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 146 /note= "Wild-type Glu substituted by Gln"
XX
XX
XX WO200227016-A2.
XX
XX 04-APR-2002.
XX
XX 25-SEP-2001; 2001WO-US29859.
XX

```

```

PR 25-SEP-2000; 2000US-235268P.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA (PATEL S B.
PA (DEAN/) DEAN M.
XX
XX Patel SB, Dean M;
XX
XX WPI: 2002-416483/44.
XX
XX Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
XX nucleic acid encoding the polypeptide, useful for treating
XX sitosterolemia, arteriosclerosis and heart diseases
XX
XX Claim 12; Page -: 66pp; English.
XX
XX The present invention relates to a new mammalian ATP-binding cassette
XX gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
XX predisposition for developing sitosterolemia, arteriosclerosis or heart
XX disease. The molecules of the invention are also useful for identifying
XX a compound which alters ABCG5 activity level comprising contacting a cell
XX culture or mammal which have ABCG5 polypeptide with a compound and
XX measuring ABCG5 biological activity in the cell culture or in mammal,
XX where an increase or decrease in ABCG5 biological activity compared to
XX ABCG5 biological activity in a control cell culture or mammal not
XX contacted with the compound, identifies a compound that increases or
XX decreases ABCG5 activity respectively. The cell culture or mammal
XX comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
XX ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
XX polypeptide in a cell culture or mammal is also compared with that of a
XX second cell culture or mammal comprising a wild type ABCG5 polypeptide.
XX stimulation of ABCG5 activity is useful for treating or preventing
XX hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
XX disease. The method of the invention is useful for increasing cholesterol
XX excretion and/or decreasing cholesterol adsorption. The present amino
XX acid sequence represents the human ABCG5 mutant E146Q protein of the
XX invention.
XX Note: This sequence is not shown in the specification but is derived
XX from the wild type human ABCG5 protein (AA096984) given on pages 35-36
XX of the specification.
XX
XX Sequence 651 AA:
XX
XX Query Match 99.98; Score 3323; DB 23; Length 651;
XX Best Local Similarity 99.88; Pred. No. 0;
XX Matches 650; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 MGLSSLTGPGSMGLQVNRGSSQLSGCAPATAPRPHSIGILHASYSVSHRVRPMMDITSC 60
Db 1 MGLSSLTGPGSMGLQVNRGSSQLSGCAPATAPRPHSIGILHASYSVSHRVRPMMDITSC 60
QY 61 RQOMTROIILKDVSLVYESGQIMCIIIGSSGSKTTLDDAMSGRLGAGTFLGEVYVNGRAL 120
Db 61 RQOMTROIILKDVSLVYESGQIMCIIIGSSGSKTTLDDAMSGRLGAGTFLGEVYVNGRAL 120
QY 121 RRQFODCFSSVYVQSDTLSSLTAVRETHTYTTALAIRGNPGSFQKVEAVMAELSLSHV 180
Db 121 RRQFODCFSSVYVQSDTLSSLTAVRETHTYTTALAIRGNPGSFQKVEAVMAELSLSHV 180
QY 181 ADRLIGNYSLGISTGERRRVSTIAAQLDPPKVMLEPDEPTGIDCMANOIVLVLELAR 240
Db 181 ADRLIGNYSLGISTGERRRVSTIAAQLDPPKVMLEPDEPTGIDCMANOIVLVLELAR 240
QY 241 RNRIVVLTIHQPSSELFQFLDKITAILSFGLIFCGTFAEMLDFFNDCGYPCEHSNPDEF 300
Db 241 RNRIVVLTIHQPSSELFQFLDKITAILSFGLIFCGTFAEMLDFFNDCGYPCEHSNPDEF 300
QY 301 YMDLTSVDIOSKEREIETSKRVOMIESAVKSAICHKTLKNIEMKHLKTLPMVPFKTKD 360
Db 301 YMDLTSVDIOSKEREIETSKRVOMIESAVKSAICHKTLKNIEMKHLKTLPMVPFKTKD 360
QY 361 SPGVFSKLGVLRRVTRNLVNRKLAIVTTRLLQNLIMGLFELFVLRVRSNVLKGAIDRV 420

```

Db 361 SPGVSKGLVLLRRVTRNLVKNKLAIVITRLQNLIMGLFLEFVLKRVSNVLKGAIDRV 420
 QY 421 GLIYQVGAATPYTGMLNANVLPVLRVAVSDQSDGLYQKQOMLAVALHVLPSVATM 480
 Db 421 GLIYQVGAATPYTGMLNANVLPVLRVAVSDQSDGLYQKQOMLAVALHVLPSVATM 480
 QY 481 IFSSVCYWTGLHPEVARFGYFSALLAPHLIGEFLLVLLGIYQNPINVSVALLSIA 540
 Db 481 IFSSVCYWTGLHPEVARFGYFSALLAPHLIGEFLLVLLGIYQNPINVSVALLSIA 540
 QY 541 GLVSGSGLRNIOEMPIPKIISYTFQKCYSEILVNEFGNLTGSSNVSSTTNPMC 600
 Db 541 GLVSGSGLRNIOEMPIPKIISYTFQKCYSEILVNEFGNLTGSSNVSSTTNPMC 600
 QY 601 AFTGQIOFIKTCPCGATSRFTMNLILYSFIPALVILGIYVFKIRDLISR 651
 Db 601 AFTGQIOFIKTCPCGATSRFTMNLILYSFIPALVILGIYVFKIRDLISR 651

RESULT 4
 AA096989 standard; Protein: 651 AA.
 AC AA096989;
 DT 30-JUL-2002 (first entry)
 DE Human ABCG5 mutant R419H protein sequence.
 KW Human: ABCG5: ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KM arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 XX mutant; muteln.
 OS Homo sapiens.
 OS Synthetic.
 XX Key Location/Qualifiers
 FT Misc-difference 419 /note= "Wild-type Arg substituted by His"
 FT MO200227016-A2.
 XX 04-APR-2002.
 PD 25-SEP-2001; 2001MO-US29859.
 PF 25-SEP-2000; 2000US-235268P.
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 XX Patel SB, Dean M;
 PI WPI: 2002-416483/44.
 DR Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 XX Claim 9; Page -: 66pp; English.
 PS The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal

CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the human ABCG5 mutant R419H protein of the
 CC invention.
 CC Note: This sequence is not shown in the specification but is derived
 CC from the wild-type human ABCG5 protein (AA096984) given on pages 35-36
 CC of the specification.
 CC Sequence 651 AA;
 SO Query Match 99.8%; Score 3321; DB 23; Length 651;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 650; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGDLSLTPGSGMGLQVNRGSSSLGAPATAPRPHSGILHASYSVSHRVRPMMDTSC 60
 Db 1 MGDLSLTPGSGMGLQVNRGSSSLGAPATAPRPHSGILHASYSVSHRVRPMMDTSC 60
 QY 61 RQWTRQILKDVSLVYESGQIMCIIIGSSSGKTTLLDMSGRLRAGFTLGEVYNGRAL 120
 Db 61 RQWTRQILKDVSLVYESGQIMCIIIGSSSGKTTLLDMSGRLRAGFTLGEVYNGRAL 120
 QY 121 RREOFQDCFSYVLQSDTLSSLVRETLHYTALLAIRGNDSFQKKEAVMAELSLSHV 180
 Db 121 RREOFQDCFSYVLQSDTLSSLVRETLHYTALLAIRGNDSFQKKEAVMAELSLSHV 180
 QY 181 ADRLIGNSLIGISIGERRRYSIAQQLQDPKVMLEFDEPTGLDQMTANOIVLLVELAR 240
 Db 181 ADRLIGNSLIGISIGERRRYSIAQQLQDPKVMLEFDEPTGLDQMTANOIVLLVELAR 240
 QY 241 RNRIVYLTIHQPSSELPQDKIAILSPGELIFCGTPREMDFNDGYPCEPSNPDEF 300
 Db 241 RNRIVYLTIHQPSSELPQDKIAILSPGELIFCGTPREMDFNDGYPCEPSNPDEF 300
 QY 301 YMDLTSVDTSKREKEIETSKRVOMIESAYKKAICHKTAKNIERNKHLKTLPMVPFKTKD 360
 Db 301 YMDLTSVDTSKREKEIETSKRVOMIESAYKKAICHKTAKNIERNKHLKTLPMVPFKTKD 360
 QY 361 SPGVSKGLVLLRRVTRNLVKNKLAIVITRLQNLIMGLFLEFVLKRVSNVLKGAIDRV 420
 Db 361 SPGVSKGLVLLRRVTRNLVKNKLAIVITRLQNLIMGLFLEFVLKRVSNVLKGAIDRV 420
 QY 421 GLIYQVGAATPYTGMLNANVLPVLRVAVSDQSDGLYQKQOMLAVALHVLPSVATM 480
 Db 421 GLIYQVGAATPYTGMLNANVLPVLRVAVSDQSDGLYQKQOMLAVALHVLPSVATM 480
 QY 481 IFSSVCYWTGLHPEVARFGYFSALLAPHLIGEFLLVLLGIYQNPINVSVALLSIA 540
 Db 481 IFSSVCYWTGLHPEVARFGYFSALLAPHLIGEFLLVLLGIYQNPINVSVALLSIA 540
 QY 541 GLVSGSGLRNIOEMPIPKIISYTFQKCYSEILVNEFGNLTGSSNVSSTTNPMC 600
 Db 541 GLVSGSGLRNIOEMPIPKIISYTFQKCYSEILVNEFGNLTGSSNVSSTTNPMC 600
 QY 601 AFTGQIOFIKTCPCGATSRFTMNLILYSFIPALVILGIYVFKIRDLISR 651
 Db 601 AFTGQIOFIKTCPCGATSRFTMNLILYSFIPALVILGIYVFKIRDLISR 651

RESULT 5
 AA096990 standard; Protein: 651 AA.
 AC AA096990;
 DT 30-JUL-2002 (first entry)

DE Human ABCG5 mutant R389H protein sequence.
XX
KW Human: ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
mutant; mutelin.
XX
OS Homo sapiens.
XX Synthetic.
FH Key Location/Qualifiers
FT Misc-difference 389 /note= "Wild-type Arg substituted by His"
XX
XX WO200227016-A2.
XX
XX 04-APR-2002.
XX
XX 25-SEP-2001; 2001WO-US29859.
XX
XX 25-SEP-2000; 2000US-235268P.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX (PATE/) PATEL S B.
XX (DEAN/) DEAN M.
XX
XX Patel SB, Dean M;
XX
XX WPI: 2002-416483/44.
XX
XX Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
PT nucleic acid encoding the polypeptide, useful for treating
PT sitosterolemia, arteriosclerosis and heart diseases
XX
XX
XX Claim 7: Page -: 66pp; English.
XX
XX The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying
CC a compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present amino
CC acid sequence represents the human ABCG5 mutant R389H protein of the
CC invention.
CC Note: This sequence is not shown in the specification but is derived
CC from the wild-type human ABCG5 protein (AA096984) given on pages 35-36
CC of the specification.
XX
XX Sequence 651 AA:
SQ

Query Match 99.8%; Score 3321; DB 23; Length 651.
Best Local Similarity 99.8%; Pred. No. 0;
Matches 650; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGDLSLTPGGSMGLQVNRGSSQLGAPATAPRPHSGILHASYSHRYRPMWDTISC 60
DB 1 MGDLSLTPGGSMGLQVNRGSSQLGAPATAPRPHSGILHASYSHRYRPMWDTISC 60
QY 61 RQMTROLKLVSLYVESGQIMCTILGSSGSKTTLIDAMSRLGAGTFLGEVYVNGRAL 120
DB 61 RQMTROLKLVSLYVESGQIMCTILGSSGSKTTLIDAMSRLGAGTFLGEVYVNGRAL 120

QY 121 RREOFODCFSYVLOSPTLLSLVRETLYHTALLAIRGNPGSFQKKVAVNAELSLSHV 180
DB 121 RREOFODCFSYVLOSPTLLSLVRETLYHTALLAIRGNPGSFQKKVAVNAELSLSHV 180
QY 181 ADRLIGNYSILGGISTGERRRVSTAAQLDOPKVMLEDEPTTGIDCANTANOIVLLVELAR 240
DB 181 ADRLIGNYSILGGISTGERRRVSTAAQLDOPKVMLEDEPTTGIDCANTANOIVLLVELAR 240
QY 241 RNRIVVLTTHORSELEFQDLKTAIISFEBLLFCGTPAEMLDPEFNDGYPCEPHSNPDE 300
DB 241 RNRIVVLTTHORSELEFQDLKTAIISFEBLLFCGTPAEMLDPEFNDGYPCEPHSNPDE 300
QY 301 YMDLTSVDPQSKERELETSEKRVOMIESAKKSKNCHTKTAKNIERMKHLTLTPVPKTKD 360
DB 301 YMDLTSVDPQSKERELETSEKRVOMIESAKKSKNCHTKTAKNIERMKHLTLTPVPKTKD 360
QY 361 SPGEVSKLGVLLRRVTRNLVRNKLAVITRLQNLIMGFLFELFVLRVRSNVLKGAIQDHY 420
DB 361 SPGEVSKLGVLLRRVTRNLVRNKLAVITRLQNLIMGFLFELFVLRVRSNVLKGAIQDHY 420
QY 421 GLLYOFVGATPYTGMLNANVLPVLRVAVSDQSDGLYQKQWQMLAYALHVLPEFSVATM 480
DB 421 GLLYOFVGATPYTGMLNANVLPVLRVAVSDQSDGLYQKQWQMLAYALHVLPEFSVATM 480
QY 481 IFSSVCYWTGLGHPFVARPGYFSAALLAPHLIGELTLVLGIVONPRTVNSVVALLSIA 540
DB 481 IFSSVCYWTGLGHPFVARPGYFSAALLAPHLIGELTLVLGIVONPRTVNSVVALLSIA 540
QY 541 GVLVSGSEFLRNIOENPDIPEKIISYTFQKCEIILVNVNFEYGLNFTCGSSNVSVTNPMC 600
DB 541 GVLVSGSEFLRNIOENPDIPEKIISYTFQKCEIILVNVNFEYGLNFTCGSSNVSVTNPMC 600
QY 601 AFTGCIQIEFKTCPCATSRFTNMFLLVSLFIPALVILGIVPKIRIDHLISR 651
DB 601 AFTGCIQIEFKTCPCATSRFTNMFLLVSLFIPALVILGIVPKIRIDHLISR 651

RESULT 6
AA096993
ID AA096993 standard; Protein; 651 AA.
XX
XX AA096993;
AC
XX
XX 30-JUL-2002 (first entry)
XX
XX Human ABCG5 mutant R419p protein sequence.
XX
XX Human: ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
mutant; mutelin.
XX
XX Homo sapiens.
XX Synthetic.
FH Key Location/Qualifiers
FT Misc-difference 419 /note= "Wild-type Arg substituted by Pro"
XX
XX WO200227016-A2.
XX
XX 04-APR-2002.
XX
XX 25-SEP-2001; 2001WO-US29859.
XX
XX 25-SEP-2000; 2000US-235268P.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX (PATE/) PATEL S B.
XX (DEAN/) DEAN M.
XX
XX Patel SB, Dean M;
XX

CC derived from mouse SSG protein referred as SEQ ID NO: 1 (AAE13289) and
CC shown in figure 7 of the specification.

XX Sequence 652 AA:

Query Match 82.6%; Score 2748.5; DB 23; Length 652;
Best Local Similarity 80.4%; Pred. No. 1.1e-280;
Matches 524; Conservative 64; Mismatches 63; Indels 1; Gaps 1;

```
OY 1 MGDLSLTPGGSMGLQVNRGSSSLEGAPATAPAP-SSLGIILHASYSVSHRVRPMDITS 59
DB 1 MGELFLPSPEGARGPHINRGSLSLEQSVTGTETARHSIGVLHVSYSVSNRGPMWNITKS 60
OY 60 CROQWTRQIILNDVSLYVESGOIMCILLSSGSGKTTLLDAMSGRGLRACFTPLGEVYVNGRA 119
DB 61 CQKMDROIILNDVSLYVESGOIMCILLSSGSGKTTLLDAMSGRGLRACFTPLGEVYVNGCE 120
OY 120 LRROFQDCFSYVLOSDFVLSLTVRETLRYTAMALCRSSADFNKRYEAVMELSLSH 179
DB 121 LRROFQDCFSYVLOSDFVLSLTVRETLRYTAMALCRSSADFNKRYEAVMELSLSH 180
OY 180 VADRILGNYSLSGISTGERRRVSTAQIILDPKVMLEPDEPTGIDCKMANOIVYLVLELA 239
DB 181 VADQMGISYNGISGERRRVSIAQIILDPKVMLEPDEPTGIDCKMANOIVYLVLELA 240
OY 240 RNRRIIVLTTHQPSSELPQFDKTAIISFGELIFCGTPAEMLDFFNDGYPCEPHSNPFD 299
DB 241 RDRRIIVYTIHQPSSELPQFDKTAIISFGELIFCGTPAEMLDFFNDGYPCEPHSNPFD 300
OY 300 FYMDLTSVDQSKERELETYSKRVQMIESAYKKSALCHTKLNTERMKHLKTLPMVPFRTK 359
DB 301 FYMDLTSVDQSKERELETYSKRVQMIESAYKKSALCHTKLNTERMKHLKTLPMVPFRTK 360
OY 360 DSPGVFSKLGVLRRVTRNLVNRKLAIVTTRLLQNLIMGLFLFVLVLRVRSNVLKAIDOR 419
DB 361 DPGMFGKLGVLRRVTRNLVNRKLAIVTTRLLQNLIMGLFLFVLVLRVRSNVLKAIDOR 420
OY 420 VGLLYQFVGATPYTGMLNAVLPFLRAVSDQESDGLYKQKMMALVALVLPFSVAT 479
DB 421 VGLLYQFVGATPYTGMLNAVLPFLRAVSDQESDGLYKQKMMALVALVLPFSVAT 480
OY 480 MIFSSVCTWTGLHPEVARFGYSFSAIILAPHLIGEFLLVLLGIYQNPINVSVALLSI 539
DB 481 VIFSSVCTWTGLHPEVARFGYSFSAIILAPHLIGEFLLVLLGIYQNPINVSVALLSI 540
OY 540 AGVLVSGSEFLRNIDEMPIFKIISFTYFQKCSSELYVNEFYGLNFTCGSSNVSTTPM 599
DB 541 SGLLIGSGFIRNIDEMPIFKIISFTYFQKCSSELYVNEFYGLNFTCGSSNVSTTPM 600
OY 600 CAFTQGIQFIKTCGATSRFTMFLIXSFIPALVILGIYVFKIRDLISR 651
DB 601 CAFTQGIQFIKTCGATSRFTMFLIXSFIPALVILGIYVFKIRDLISR 652
```

RESULT 8

AAE13289
ID AAE13289 standard; Protein; 652 AA.

XX AAE13289;

XX 12-FEB-2002 (first entry)

XX Mouse sitosterolemia susceptibility gene (SSG) protein.

XX Mouse; sitosterolemia susceptibility gene; SSG; atherosclerosis;
XX sterol-related disorder; hyperlipidaemia; hypercholesterolaemia;
XX gall stone; coronary heart disease; cardiovascular disease; arthritis;
XX xanthoma; haemolytic anaemia; transgenic animal; chromosome 17; therapy.

OS Mus sp.

XX WO200179272-A2.

PD 25-OCT-2001.

XX 18-APR-2001; 2001WO-US12758.

XX 18-APR-2000; 2000US-198465P.

XX 15-MAY-2000; 2000US-204234P.

XX (TULIA-) TULARK INC.

XX Tian H, Schultz J, Shan B;

XX WPI; 2002-017598/02.

XX N-PSDB; AAD22008.

PT Novel sitosterolemia susceptibility gene polypeptide and
PT polynucleotide, useful for screening a compound that increases the
PT level of expression or activity of SSG polypeptide for treating
PT sterol-related disorder

PS Claim 19; Fig 7; 105pp; English.

CC The invention relates to an isolated sitosterolemia susceptibility gene
CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
CC binding cassette (ABC) family cholesterol transporter. SSG is useful
CC for identifying a compound useful in the treatment or prevention of a
CC sterol-related disorder, including sitosterolemia, hyperlipidaemia,
CC hypercholesterolemia, gall stones, HDL deficiency, atherosclerosis or
CC nutritional deficiencies. SSG is also useful for treating cholesterol-
CC associated diseases or conditions including coronary heart disease and
CC other cardiovascular diseases, and sitosterolemia-associated condition
CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
CC expression cassette is useful in the production of transgenic non-human
CC animals. SSG genes and their homologues are useful as tools for a number
CC of applications including diagnosing sitosterolemia and other
CC cardiovascular disorders, for forensics and paternity determinations,
CC and for treating any of a large number of SSG associated diseases. The
CC present sequence is mouse SSG protein. Mouse SSG is located on
CC chromosome 17.

XX Sequence 652 AA:

Query Match 82.5%; Score 2744.5; DB 23; Length 652;
Best Local Similarity 80.2%; Pred. No. 2.9e-280;
Matches 523; Conservative 64; Mismatches 64; Indels 1; Gaps 1;

```
OY 1 MGDLSLTPGGSMGLQVNRGSSSLEGAPATAPAP-SSLGIILHASYSVSHRVRPMDITS 59
DB 1 MGELFLPSPEGARGPHINRGSLSLEQSVTGTETARHSIGVLHVSYSVSNRGPMWNITKS 60
OY 60 CROQWTRQIILNDVSLYVESGOIMCILLSSGSGKTTLLDAMSGRGLRACFTPLGEVYVNGRA 119
DB 61 CQKMDROIILNDVSLYVESGOIMCILLSSGSGKTTLLDAMSGRGLRACFTPLGEVYVNGCE 120
OY 120 LRROFQDCFSYVLOSDFVLSLTVRETLRYTAMALCRSSADFNKRYEAVMELSLSH 179
DB 121 LRROFQDCFSYVLOSDFVLSLTVRETLRYTAMALCRSSADFNKRYEAVMELSLSH 180
OY 180 VADRILGNYSLSGISTGERRRVSTAQIILDPKVMLEPDEPTGIDCKMANOIVYLVLELA 239
DB 181 VADQMGISYNGISGERRRVSIAQIILDPKVMLEPDEPTGIDCKMANOIVYLVLELA 240
OY 240 RNRRIIVLTTHQPSSELPQFDKTAIISFGELIFCGTPAEMLDFFNDGYPCEPHSNPFD 299
DB 241 RDRRIIVYTIHQPSSELPQFDKTAIISFGELIFCGTPAEMLDFFNDGYPCEPHSNPFD 300
OY 300 FYMDLTSVDQSKERELETYSKRVQMIESAYKKSALCHTKLNTERMKHLKTLPMVPFRTK 359
DB 301 FYMDLTSVDQSKERELETYSKRVQMIESAYKKSALCHTKLNTERMKHLKTLPMVPFRTK 360
OY 360 DSPGVFSKLGVLRRVTRNLVNRKLAIVTTRLLQNLIMGLFLFVLVLRVRSNVLKAIDOR 419
DB 361 DPGMFGKLGVLRRVTRNLVNRKLAIVTTRLLQNLIMGLFLFVLVLRVRSNVLKAIDOR 420
```


OY 420 VGLLYQFVGATPYTGMLNANVLPVLRVAVSDQSDGLYOKWOMMLAYALVLPESVAT 479
 CC 421 VGLLYQFVGATPYTGMLNANVLPVLRVAVSDQSDGLYOKWOMMLAYALVLPESVAT 480
 Db 421 VGLLYQFVGATPYTGMLNANVLPVLRVAVSDQSDGLYOKWOMMLAYALVLPESVAT 480
 OY 480 MIFSSVCYWTGLGHPVAREGYFSALLAPHLIGEFLLVLLGIYONPNIVNSVALLSI 539
 CC 481 VIFSSVCYWTGLGHPVAREGYFSALLAPHLIGEFLLVLLGIYONPNIVNSVALLSI 540
 Db 481 VIFSSVCYWTGLGHPVAREGYFSALLAPHLIGEFLLVLLGIYONPNIVNSVALLSI 540
 OY 540 AGVLVSGFLRNIOEMPPIPKIISYFFQKCYSEILVNEFYGLNFTCGSSNVSTNPM 599
 CC 541 SGLLIGSGFIRNIQEMPIPKIIGYFFQKCYSEILVNEFYGLNFTCGSSNVSTNPM 600
 Db 541 SGLLIGSGFIRNIQEMPIPKIIGYFFQKCYSEILVNEFYGLNFTCGSSNVSTNPM 600
 OY 600 CAFTGOGIOFIKTCPCGATSRFTNMFLIYFTIPALVILGIYVKIRIDHLISR 651
 CC 601 CAITGOGVQFIKTCPCGATSRFTNMFLIYFTIPALVILGIYVKIRIDHLISR 652
 Db 601 CAITGOGVQFIKTCPCGATSRFTNMFLIYFTIPALVILGIYVKIRIDHLISR 652
 RESULT 9
 AAE13308
 ID AAE13308 standard; Protein; 652 AA.
 XX
 AC AAE13308;
 XX
 DT 12-FEB-2002 (first entry)
 DE Mouse sitosterolemia susceptibility gene (SSG) protein variant #1.
 XX
 KW Mouse; sitosterolemia susceptibility gene; SSG; atherosclerosis; muten;
 KM sterol-related disorder; hyperlipidemia; hypercholesterolemia; mutant;
 KM gall stone; coronary heart disease; cardiovascular disease; arthritis;
 KM xanthoma; haemolytic anaemia; transgenic animal; therapy; variant.
 XX
 OS Mus sp.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 17 /note="Wild type ile substituted with leu"
 XX
 PN W0200179272-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 18-APR-2001; 2001MO-US12758.
 XX
 PR 18-APR-2000; 2000US-198465P.
 PR 15-MAY-2000; 2000US-204234P.
 XX
 PA (TDLA-) TULARIK INC.
 XX
 PI Tlan H, Schultz J, Shan B;
 XX
 DR WPI; 2002-017598/02.
 XX
 PT Novel sitosterolemia susceptibility gene polypeptide and
 PT polynucleotide, useful for screening a compound that increases the
 PT level of expression or activity of SSG polypeptide for treating
 PT sterol-related disorder
 XX
 PS Disclosure: Page -: 105pp; English.
 XX
 CC The invention relates to an isolated sitosterolemia Susceptibility Gene
 CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
 CC binding cassette (ABC) family cholesterol transporter. SSG is useful
 CC for identifying a compound useful in the treatment or prevention of a
 CC sterol-related disorder, including sitosterolemia, hyperlipidaemia,
 CC hypercholesterolemia, gall stones, HDL deficiency, atherosclerosis or
 CC nutritional deficiencies. SSG is also useful for treating cholesterol-
 CC associated diseases or conditions including coronary heart disease and
 CC other cardiovascular diseases, and sitosterolemia-associated condition
 CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
 CC expression cassette is useful in the production of transgenic non-human
 CC animals. SSG genes and their homologues are useful as tools for a number

CC of applications including diagnosing sitosterolemia and other
 CC cardiovascular disorders, for forensics and paternity determinations,
 CC and for treating any of a large number of SSG associated diseases. The
 CC present sequence is mouse SSG protein variant obtained by replacing
 CC Ile17 with Leu.
 CC Note: The present sequence is not shown in the specification but is
 CC derived from mouse SSG protein referred as SEQ ID NO: 1 (AAE13289) and
 CC shown in figure 7 of the specification.
 XX
 SQ Sequence 652 AA;
 Query Match 82.5%; Score 2742.5; DB 23; Length 652;
 Best Local Similarity 80.2%; Pred. No. 4,76-280;
 Matches 523; Conservative 64; Mismatches 64; Indels 1; Gaps 1;
 OY 1 MGDLSLTPGSGMGLQVNRGSSGLECAPATAPP-HSLGILNASYSVSHRPMWDITS 59
 Db 1 MGDLSLTPGSGMGLQVNRGSSGLECAPATAPP-HSLGILNASYSVSHRPMWDITS 59
 OY 60 CROQWROIILKDVSLYVESGOIMCILLSSSGSKTLLDAMSGRLGRAGTFLGEYVNGRA 119
 Db 60 CROQWROIILKDVSLYVESGOIMCILLSSSGSKTLLDAMSGRLGRAGTFLGEYVNGRA 119
 OY 61 CQKMDROIILKDVSLYVESGOIMCILLSSSGSKTLLDAMSGRLGRAGTFLGEYVNGCE 120
 Db 61 CQKMDROIILKDVSLYVESGOIMCILLSSSGSKTLLDAMSGRLGRAGTFLGEYVNGCE 120
 OY 120 LRREOFDCESYVLSQSDTLSSLVRETLHYTALLAIRGNPGSGFQKKEVAVMAELSLH 179
 Db 120 LRREOFDCESYVLSQSDTLSSLVRETLHYTALLAIRGNPGSGFQKKEVAVMAELSLH 179
 OY 121 LRROQFODCSYVLSQSDTLSSLVRETLHYTALLAIRGNPGSGFQKKEVAVMAELSLH 180
 Db 121 LRROQFODCSYVLSQSDTLSSLVRETLHYTALLAIRGNPGSGFQKKEVAVMAELSLH 180
 OY 180 VADRLINYSLSGISTEERRRVSIAAQLDDPKVYLFDEPTGLDCKTANOIVLLVELA 239
 Db 180 VADRLINYSLSGISTEERRRVSIAAQLDDPKVYLFDEPTGLDCKTANOIVLLVELA 239
 OY 181 VADOMISYNGFGISSEERRRVSIAAQLDDPKVYLFDEPTGLDCKTANOIVLLVELA 240
 Db 181 VADOMISYNGFGISSEERRRVSIAAQLDDPKVYLFDEPTGLDCKTANOIVLLVELA 240
 OY 240 RNRRIIVLTTHOPSEIFOLFDRKAIISFGELIFCGTPAEMLDPEFCGCPCHSPDFD 299
 Db 240 RNRRIIVLTTHOPSEIFOLFDRKAIISFGELIFCGTPAEMLDPEFCGCPCHSPDFD 299
 OY 241 RNRRIIVLTTHOPSEIFOLFDRKAIISFGELIFCGTPAEMLDPEFCGCPCHSPDFD 300
 Db 241 RNRRIIVLTTHOPSEIFOLFDRKAIISFGELIFCGTPAEMLDPEFCGCPCHSPDFD 300
 OY 300 FYMDLTSVDTQSKERELETSKRYQMISAYKKSNAICRTLKNIEMRMLKTLPMVPKTK 359
 Db 300 FYMDLTSVDTQSKERELETSKRYQMISAYKKSNAICRTLKNIEMRMLKTLPMVPKTK 359
 OY 301 FYMDLTSVDTQSKERELETSKRYQMISAYKKSNAICRTLKNIEMRMLKTLPMVPKTK 360
 Db 301 FYMDLTSVDTQSKERELETSKRYQMISAYKKSNAICRTLKNIEMRMLKTLPMVPKTK 360
 OY 360 DSGVSEKILAVLRVTRNLYRNKLAIVTLLNLINGLFLFVLRVNSVNLKGAIDR 419
 Db 360 DSGVSEKILAVLRVTRNLYRNKLAIVTLLNLINGLFLFVLRVNSVNLKGAIDR 419
 OY 361 DPGMFKILAVLRVTRNLYRNKLAIVTLLNLINGLFLFVLRVNSVNLKGAIDR 420
 Db 361 DPGMFKILAVLRVTRNLYRNKLAIVTLLNLINGLFLFVLRVNSVNLKGAIDR 420
 OY 420 VGLLYQFVGATPYTGMLNANVLPVLRVAVSDQSDGLYOKWOMMLAYALVLPESVAT 479
 Db 420 VGLLYQFVGATPYTGMLNANVLPVLRVAVSDQSDGLYOKWOMMLAYALVLPESVAT 479
 OY 421 VGLLYQFVGATPYTGMLNANVLPVLRVAVSDQSDGLYOKWOMMLAYALVLPESVAT 480
 Db 421 VGLLYQFVGATPYTGMLNANVLPVLRVAVSDQSDGLYOKWOMMLAYALVLPESVAT 480
 OY 480 MIFSSVCYWTGLGHPVAREGYFSALLAPHLIGEFLLVLLGIYONPNIVNSVALLSI 539
 Db 480 MIFSSVCYWTGLGHPVAREGYFSALLAPHLIGEFLLVLLGIYONPNIVNSVALLSI 539
 OY 481 VIFSSVCYWTGLGHPVAREGYFSALLAPHLIGEFLLVLLGIYONPNIVNSVALLSI 540
 Db 481 VIFSSVCYWTGLGHPVAREGYFSALLAPHLIGEFLLVLLGIYONPNIVNSVALLSI 540
 OY 540 AGVLVSGFLRNIOEMPPIPKIISYFFQKCYSEILVNEFYGLNFTCGSSNVSTNPM 599
 Db 540 AGVLVSGFLRNIOEMPPIPKIISYFFQKCYSEILVNEFYGLNFTCGSSNVSTNPM 599
 OY 541 SGLLIGSGFIRNIQEMPIPKIIGYFFQKCYSEILVNEFYGLNFTCGSSNVSTNPM 600
 Db 541 SGLLIGSGFIRNIQEMPIPKIIGYFFQKCYSEILVNEFYGLNFTCGSSNVSTNPM 600
 OY 600 CAFTGOGIOFIKTCPCGATSRFTNMFLIYFTIPALVILGIYVKIRIDHLISR 651
 Db 600 CAFTGOGIOFIKTCPCGATSRFTNMFLIYFTIPALVILGIYVKIRIDHLISR 651
 OY 601 CAITGOGVQFIKTCPCGATSRFTNMFLIYFTIPALVILGIYVKIRIDHLISR 652
 Db 601 CAITGOGVQFIKTCPCGATSRFTNMFLIYFTIPALVILGIYVKIRIDHLISR 652
 RESULT 10
 AAU096985
 ID AAU096985 standard; Protein; 652 AA.
 XX
 AC AAU096985;
 XX
 DT 30-JUL-2002 (first entry)
 DE Mouse ABCG5 protein.
 XX
 KW Mouse; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KM atherosclerosis; heart disease; hypersterolemia; Alzheimer's disease.
 XX

OS Mus sp. Location/Qualifiers
 XX Key
 FH Misc-difference 638.652
 FT /note="Encoded by CTag"
 XX MO200227016-A2.
 XX 04-APR-2002.
 XX 25-SEP-2001: 2001WO-US29859.
 XX 25-SEP-2000: 2000US-235268P.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATEL S B.
 PA (DEAN/) DEAN M.
 PI Patel SB, Dean M:
 XX MPI: 2002-416483/44.
 DR N-PSDB; ABR51684.
 XX
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT strossterolemia, arteriosclerosis and heart diseases
 XX
 PS Example 3; Page 42; 66pp; English.
 XX
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing strossterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the mouse ABCG5 protein of the invention.
 XX
 SQ Sequence 652 AA:
 Query Match 82.3%; Score 2738.5; DB 23; Length 652;
 Best Local Similarity 80.1%; Pred No. 1.2e-279;
 Matches 522; Conservative 64; Mismatches 63; Indels 1; Gaps 1;
 QY 1 MGDLSLRPGSGMGLQVNGNQSOSLEGAPATAPPE-HSLGILHASYSHRYRPMMDITS 59
 DB 1 MGELPLSPGARGHNRHNSLSLEQSVTGTEARHSLGVLHVSYSNNRGPMWNKIS 60
 QY 60 CROOFTROIIDVSLYVSSGQIMCLGSSGSKTTLDDAMSGRGRACFTJGEVYVNGRA 119
 DB 61 COQKMDROIIDVSLYIESGQIMCLGSSGSKTTLDAISGRRLRTGTLEGEVYVNGCE 120
 QY 120 LRREFOFDCFSYVLSQDFLSSIFVRETLHTALIAIRKNGSGKQKVEAVMELSLSH 179
 DB 121 LRROFQOCFSYVLSQDFLSSIFVRETLHTALIAIRKNGSGKQKVEAVMELSLSH 180
 QY 180 VADRILGYSLSGISTGERRRVSIAAQLADPKVLEDEPTGDLCKMANOIVYLVLELA 239
 DB 181 VADWIGISYNGGSISSGERRRVSIAAQLADPKVLEDEPTGDLCKMANOIVYLVLELA 240
 QY 240 RNRIVLTITQPSSELPFLDKATALLSGELIFCGTPAEKMLDFPNDGYPCEHSNPF 299

DB 241 RRDRIIVITIHOPRSELPFLDKATALLSGELIFCGTPAEKMLDFPNDGYPCEHSNPF 300
 QY 300 FYMDLTSVDTOSKEREIETSKRQVOMIESAYKSAICHTLKNTERKHKHTLPMVPEKTK 359
 DB 301 FYMDLTSVDTOSKEREIETSKRQVOMIESAYKSAICHTLKNTERKHKHTLPMVPEKTK 360
 QY 360 DSEGVESKLGVLRLRVTLNRNKLAVITRLLQNLIMGLFLEFVLRVSNVNLKGAIQDR 419
 DB 361 DPGMEGKLGVLRLRVTLNRNKLAVITRLLQNLIMGLFLEFVLRVSNVNLKGAIQDR 420
 QY 420 VGLLYQVATPTGCLNANVNLFPVLRVAVSDSDGGLYOKQOMLAVLHVLPEFSVAT 479
 DB 421 VGLLYQVATPTGCLNANVNLFPVLRVAVSDSDGGLYOKQOMLAVLHVLPEFSVAT 480
 QY 480 MIFSSVCYWTGLGHPVARGYFSALAPHLIGELTTLVIGIYONRNIVNSVALLSI 539
 DB 481 VIFSSVCYWTGLGHPVARGYFSALAPHLIGELTTLVIGIYONRNIVNSVALLSI 540
 QY 540 AGVLVSGEFLRNIOEMPIPKIISYFTPOKCYSEILVNEFEGLNFTCGSSNVSTTNM 599
 DB 541 SGLIGSGFIRNIOEMPIPKIISYFTPOKCYSEILVNEFEGLNFTCGSSNVSTTNM 600
 QY 600 CAPTQGIOTIEKTCPCGATSRFTMNLILYSFIPALVILGIVVEFKIDHLISR 651
 DB 601 CAITQGVPIEKTCPGATSRFTMNLILYSFIPALVILGIVVEFKIDHLISR 652
 RESULT 11
 ID AAU96986 standard; Protein; 652 AA.
 XX AAU96986;
 AC 30-JUL-2002 (first entry)
 XX
 DE Rat ABCG5 protein.
 XX
 KW Rat: ABCG5; ATP-binding cassette gene 5; strossterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease.
 OS Rat sp.
 XX
 PN WO200227016-A2.
 XX
 PD 04-APR-2002.
 XX
 XX 25-SEP-2001: 2001WO-US29859.
 XX 25-SEP-2000: 2000US-235268P.
 XX
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATEL S B.
 PA (DEAN/) DEAN M.
 PI Patel SB, Dean M:
 XX MPI: 2002-416483/44.
 DR N-PSDB; ABR51684.
 XX
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT strossterolemia, arteriosclerosis and heart diseases
 XX
 PS Example 3; Page 45; 66pp; English.
 XX
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing strossterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,

CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the rat ABCG5 protein of the invention.

xx Sequence 652 AA;

Query Match 82.0%; Score 2727.5; DB 23; Length 652;
 Best Local Similarity 79.4%; Pred. No. 1.8e-278;
 Matches 518; Conservative 68; Mismatches 65; Indels 1; Gaps 1;

OY 1 MGDLSLTPGSGMGLQVNRGSSGLECAPATAPPE-HSLGILHASYSHRYRPMWDTSC 59
 DB 1 MGLPFLSPGAPPHNNRGSOSLSLEGSVTGSEARHSLGLVNLVSFSVSNVGPWMNKS 60
 OY 60 CROQWTRQILKDVSLYVESGOIMCIGSSGKTTLLDAMSGRLGRAGTFGEYVNGRA 119
 DB 61 CQKMDKRLKDVSLYVESGOTKCTIGSSGKTTLLDAMSGRLGRAGTFGEYVNGRA 120
 OY 120 LRREPODCFSYVLSQDTLLSLTVRETLAYTALLAIRGNPGSFQKKEAVAMELSLH 179
 DB 121 LRROFQDCFSYVLSQDVFLLSLTVRETLAYTALLAIRGNPGSFQKKEAVAMELSLH 180
 OY 180 VADLLIGNYSIGSTGRRRVSTAALQLOPKWMLPEPTTGDCMTANQIVVLVLA 239
 DB 181 VADQMTGNVNGSISGERRRVSTAALQLOPKWMLPEPTTGDCMTANQIVVLVLA 240
 OY 240 RRRRIYVLTTHQPSSELFQLEKTAISFGELIFCGFPAEMLEFFNDCGCPESHSNFD 299
 DB 241 RRRRIYVLTTHQPSSELFQLEKTAISFGELIFCGFPAEMLEFFNDCGCPESHSNFD 300
 OY 300 FYMDLTVSDIOSKRELETSKRVOMIESAYKSAICHTKTKNIEEMKHLKTLPMVPFKTK 359
 DB 301 FYMDLTVSDIOSKRELETSKRVOMIESAYKSAICHTKTKNIEEMKHLKTLPMVPFKTK 360
 OY 360 DSPGVESKLVLRVRNRLVNRKLAVTRLLQNLIMGELTFEVLPRSVNLGATIDR 419
 DB 361 NPPGFCKLVLRVRNRLVNRKLAVTRLLQNLIMGELTFEVLPRSVNLGATIDR 420
 OY 420 VGLLYQFVGATPYGMNLAVNLFPVLRAVSDQSDGLYOKMOMLAVLAVLPFSYAT 479
 DB 421 VGLLYQFVGATPYGMNLAVNLFPVLRAVSDQSDGLYOKMOMLAVLAVLPFSYAT 480
 OY 480 MISSVCYWTGLHPEVARFGYSALIALPHLIGEFLLTVLLGIVONPNIVSYVALLSI 539
 DB 481 VIFSSVCYWTGLHPEVARFGYSALIALPHLIGEFLLTVLLGIVONPNIVSYVALLSI 540
 OY 540 AGVAVSGELNIDEMPIPKIISYFTFOKCSLIVVNEGYGNTFCGSSNVSYTPPM 599
 DB 541 SGLIIGSGELNIDEMPIPKIISYFTFOKCSLIVVNEGYGNTFCGSSNVSYTPPM 600
 OY 600 CAPTOGLOFTEKTCGATSRFTNFIILYSIPALVILGIYVFKIRDLISR 651
 DB 601 CSMTOGLOFTEKTCGATSRFTNFIILYSIPALVILGIYVFKIRDLISR 652

RESULT 12

AA096991 ID AA096991 standard; Protein: 408 AA.

AC AA096991;

DT 30-JUL-2002 (first entry)

XX

DE Human ABCG5 mutant R408x protein sequence.
 XX
 KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW mutant; muten.
 OS Homo sapiens.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Misc-difference 408
 FT /note="Wild-type protein truncated at this position"

W0200227016-A2.

04-APR-2002.

25-SEP-2001: 2001WO-US29859.

25-SEP-2000; 2000US-235268P.

(USSH) US DEPT HEALTH & HUMAN SERVICES.
 (PATE/) PATEL S B.
 (DEAN/) DEAN M.

Patel SB, Dean M;

WPI; 2002-416483/44.

Novel mammalian ATP-binding cassette gene 5 polypeptide, and the

nucleic acid encoding the polypeptide, useful for treating

sitosterolemia, arteriosclerosis and heart diseases

Claim 10; Page -: 66pp; English.

The present invention relates to a new mammalian ATP-binding cassette gene 5 (ABCG5) polypeptide. The invention is useful for identifying a predisposition for developing sitosterolemia, arteriosclerosis or heart disease. The molecules of the invention are also useful for identifying a compound which alters ABCG5 activity level comprising contacting a cell culture or mammal which have ABCG5 polypeptide with a compound and measuring ABCG5 biological activity in the cell culture or in mammal, where an increase or decrease in ABCG5 biological activity compared to ABCG5 biological activity in a control cell culture or mammal not contacted with the compound, identifies a compound that increases or decreases ABCG5 activity respectively. The cell culture or mammal comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The ABCG5 biological activity, or level of ABCG5 mRNA, or level of the polypeptide in a cell culture or mammal is also compared with that of a second cell culture or mammal comprising a wild type ABCG5 polypeptide. Stimulation of ABCG5 activity is useful for treating or preventing hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's disease. The method of the invention is useful for increasing cholesterol excretion and/or decreasing cholesterol adsorption. The present amino acid sequence represents the human ABCG5 mutant R408x protein of the invention.
 Note: This sequence is not shown in the specification but is derived from the wild-type human ABCG5 protein (AA096984) given on pages 35-36 of the specification.

Sequence 408 AA;

Query Match 62.6%; Score 2081; DB 23; Length 408;
 Best Local Similarity 100.0%; Pred. No. 1.7e-210;
 Matches 408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MGDLSLTPGSGMGLQVNRGSSGLECAPATAPPEHSLGILHASYSHRYRPMWDTSC 60
 DB 1 MGDLSLTPGSGMGLQVNRGSSGLECAPATAPPEHSLGILHASYSHRYRPMWDTSC 60
 OY 61 RQOWTRQILKDVSLYVESGOIMCIGSSGKTTLLDAMSGRLGRAGTFGEYVNGRAL 120
 DB 61 RQOWTRQILKDVSLYVESGOIMCIGSSGKTTLLDAMSGRLGRAGTFGEYVNGRAL 120

QY 121 RRFQDPCFVYQSDTLLSLVRETHHTALLAIRGNPGSFQKVEAVMAELSLSHV 180
 |||||||
 DB 121 RRFQDPCFVYQSDTLLSLVRETHHTALLAIRGNPGSFQKVEAVMAELSLSHV 180
 QY 181 ADRLIGNYSIGISTGERRRVSIAAQLDPPKMLDEPTTGIDCMANQIVLVLELAR 240
 |||||||
 DB 181 ADRLIGNYSIGISTGERRRVSIAAQLDPPKMLDEPTTGIDCMANQIVLVLELAR 240
 QY 241 RNRIVVLTTHQPSSELFQLEFDKTAIISFGELIFCGPAEMLDFPNDGYPCEPHSNPDF 300
 |||||||
 DB 241 RNRIVVLTTHQPSSELFQLEFDKTAIISFGELIFCGPAEMLDFPNDGYPCEPHSNPDF 300
 QY 301 YMDLTSVDPQSKRELETETSKRVOMISAKKSAICHKTKLNIERMHLKTLPMVPPKTKD 360
 |||||||
 DB 301 YMDLTSVDPQSKRELETETSKRVOMISAKKSAICHKTKLNIERMHLKTLPMVPPKTKD 360
 QY 361 SPQVFSKGLVLRRTVRNLRVNRKLAVITRLQNLINGLFLFVLVR 408
 |||||||
 DB 361 SPQVFSKGLVLRRTVRNLRVNRKLAVITRLQNLINGLFLFVLVR 408

RESULT 13

AA096987

ID AA096987 standard; Protein; 340 AA.

AC AA096987;

DT 30-JUL-2002 (first entry)

DE Hamster ABCG5 protein.

KM Hamster: ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease.

OS *Cricetinae* sp.

PN W0200227016-A2.

PD 04-APR-2002.

PD 25-SEP-2001; 2001WO-US29859.

PR 25-SEP-2000; 2000US-235268P.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES;
 PA (PATEL S B.
 PA (DEAN/) DEAN M.

PI Patel SB, Dean M;
 DR WPI: 2002-416483/44.
 DR N-PSDB; ABR51687.

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 PS Example 3; Page 46; 66pp; English.

CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the

CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present amino
 CC acid sequence represents the hamster ABCG5 protein of the invention.

SQ Sequence 340 AA:

Query Match 41.7%; Score 1387.5; DB 23; Length 340;
 best Local Similarity 78.6%; Pred. No. 2,2e-137;
 Matches 264; Conservative 39; Mismatches 32; Indels 1; Gaps 1;

QY 98 AMSGRGRAGTFLGEYVNGRALRRQDPCFVYQSDTLLSLVRETHHTALLAIR 157
 1 AISGRRLRRGTLEGEYVNGRELRRQDPCFVYQSDTLLSLVRETHHTALLAIR 60
 DB 158 RGNPGSFQKVEAVMAELSLSHVADRLIGNYSIGISTGERRRVSIAAQLDPPKMLFD 217
 61 SSSSDFYDKKVEAVMAELSLSHVADRLIGNYSIGISTGERRRVSIAAQLDPPKMLMD 120
 QY 218 EPTTGIDCMANQIVLVLELARNRIVVLTTHQPSSELFQLEFDKTAIISFGELIFCGP 277
 |||||||
 DB 121 EPTTGIDCMANQIVLVLELARNRIVVLTTHQPSSELFQLEFDKTAIISFGELIFCGP 180
 QY 278 AEMLDFPNDGYPCEPHSNPDFYMDLTSVDPQSKRELETETSKRVOMISAKKSAICHK 337
 |||||||
 DB 181 EEMLDFPNDGYPCEPHSNPDFYMDLTSVDPQSKRELETETSKRVOMISAKKSAICHK 240
 QY 338 TLKNIERMHLKTLPMVPPKTKDSPQVFSKGLVLRRTVRNLRVNRKLAVITRLQNLING 397
 |||||||
 DB 241 TLKNIERMHLKTLPMVPPKTKDSPQVFSKGLVLRRTVRNLRVNRKLAVITRLQNLING 300
 QY 398 LFLFFVLVRNLRVNRKLAVITRLQNLINGLFLFVLVR 433
 |||||||
 DB 301 LFLFFVLVRNLRVNRKLAVITRLQNLINGLFLFVLVR 335

RESULT 14

AA096988

ID AA096988 standard; Protein; 243 AA.

AC AA096988;

DT 30-JUL-2002 (first entry)

DE Human ABCG5 mutant R243X protein sequence.

KM Human: ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 mutant; mutlein.

OS Homo sapiens.
 OS Synthetic.

FT Key Location/Qualifiers
 FT Misc-difference 243 /note= "Wild-type protein truncated at this position"

W0200227016-A2.

PD 04-APR-2002.

PD 25-SEP-2001; 2001WO-US29859.

PR 25-SEP-2000; 2000US-235268P.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES;
 PA (PATEL S B.
 PA (DEAN/) DEAN M.

PI Patel SB, Dean M;
 DR WPI: 2002-416483/44.
 DR N-PSDB; ABR51687.

XX WPI; 2002-416483/44.
DR
XX
PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
PT nucleic acid encoding the polypeptide, useful for treating
PT sitosterolemia, arteriosclerosis and heart diseases
XX
PS Claim 13; Page -: 66pp; English.
XX
CC The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying
CC a compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present amino
CC acid sequence represents the human ABCG5 mutant R243X protein of the
CC invention.
CC Note: This sequence is not shown in the specification but is derived
CC from the wild-type human ABCG5 protein (AA96984) given on pages 35-36
CC of the specification..
XX
SQ Sequence 243 AA:
Query Match 37 1%; Score 1234; DB 23; Length 243;
Best Local Similarity 100.0%; Pred. No. 2.1e-121;
Matches 243; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MGDLSLTTPGSGMGQVNRGSSSSLEGAPATAPPEPHSGITLHASVSVSHRVRPMWDTSC 60
DB 1 MGDLSLTTPGSGMGQVNRGSSSSLEGAPATAPPEPHSGITLHASVSVSHRVRPMWDTSC 60
OY 61 RQOMTRQILKDVSLVSGQIMCJLSSGSGKTTLLDMSGRLAGAGTFLEVVYNGRAL 120
DB 61 RQOMTRQILKDVSLVSGQIMCJLSSGSGKTTLLDMSGRLAGAGTFLEVVYNGRAL 120
OY 121 RREQDQDFSVYLOSDTLLSSTTRETHTYALLAIRGNGSFQKKEAVMAELSLSHV 180
DB 121 RREQDQDFSVYLOSDTLLSSTTRETHTYALLAIRGNGSFQKKEAVMAELSLSHV 180
OY 181 ADRLIGNYSLGISSTGERRRYSIAQQLQDPKVMFLDEPTGLDQMTANOIVLLVELAR 240
DB 181 ADRLIGNYSLGISSTGERRRYSIAQQLQDPKVMFLDEPTGLDQMTANOIVLLVELAR 240
OY 241 RNR 243
DB 241 RNR 243
RESULT 15
AAB41856
ID AAB41856 standard; Protein; 144 AA.
XX
AC AAB41856;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human OREF1620 polypeptide sequence SEQ ID NO:3240.
XX
KW Human; open reading frame; OREF; detection; cytostatic; hepatotropic;
KW vulnary; antipariatic; antiparkinsonian; nootropic; neuroprotective;

KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antineumatic; antihypoid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; nocturnal haemoglobinuria; burn; wound;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
XX
PR 02-APR-1999; 99US-0127636.
XX
PR 05-APR-1999; 99US-0127728.
XX
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CBRAGEN CORP.
XX
PI Shinkets PA, Leach M;
XX
DR WPI; 2000-602362/57.
XX
DR N-PSDB; AAC76065.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 11; Page 2444; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human OREF open reading frames 1 to 3161. The OREF
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antipariatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antihypoid; and antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an OREF-associated disorder. The
CC nucleic acids can be used to express OREF proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 144 AA:
Query Match 21.7%; Score 722; DB 21; Length 144;
Best Local Similarity 99.3%; Pred. No. 1.1e-67;
Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 357 KTKDSPGVFSKGLVLRVTRNLTAVRNLAVITRLQNLGLFLFVLRVRSVVLKAI 416
DB 1 KTKDSPGVFSKGLVLRVTRNLTAVRNLAVITRLQNLGLFLFVLRVRSVVLKAI 60
OY 417 QDRVGLLYQFVGATPYTGMLNAVLPVLRVAVSDQSDGLYQKWQMLLYALHVLVPSV 476
|||||

Mon Jul 28 09:43:29 2003

us-09-989-981a-6.rag

Page 13

Db	61	ODRVGLYQFVGATPYTGMNANVNFVLRVAVSDQSDGLYOKKOMMLAYALHVLPPSV	120
OY	477	VATMIFSSVCYWTGLHPEVARFG	500
Db	121	VATMIFSSVCYWTGLHPEVARLIG	144

Search completed: June 11, 2003, 09:12:15
Job time : 43 secs

OM nucleic - nucleic search, using sw model

Run on: July 26, 2003, 23:50:29 ; Search time 4351.59 Seconds
(without alignments)
13069.363 Million cell updates/sec

Title:	US-09-989-981A-5
Perfect score:	3240

Sequence: 1 gtcaggtgagcagcgagg.....aatattcataacatctgg 2340

Scoring table: IDENTITY_NDC .
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1:	em_estha:*
2:	em_esthum:*
3:	em_estin:*
4:	em_estnu:*
5:	em_estov:*
6:	em_estpl:*
7:	em_estro:*
8:	em_hc:*
9:	gb_esti:*
10:	gb_est2:*
11:	gb_hc:*
12:	gb_est3:*
13:	gb_est4:*
14:	gb_est5:*
15:	em_estfun:*
16:	em_estom:*
17:	em_gss_inu:*
18:	em_gss_inu1:*
19:	em_gss_pln:*
20:	em_gss_vrt:*
21:	em_gss_fun:*
22:	em_gss_man:*
23:	em_gss_mus:*
24:	em_gss_pro:*
25:	em_gss_pod:*
26:	em_gss_phg:*
27:	em_gss_vrl:*
28:	gb_gss1:*
29:	gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	544	23.2	594	9	AV689089	AV689089 AV689089
2	521	22.3	597	9	AV694671	AV694671 AV694671
3	477	20.4	477	9	AV720911	AV720911 AV720911
4	463.2	19.8	713	10	BB598373	BB598373 BB598373

C	5	429	18.3	432	9	AT033358
C	6	418	17.9	418	9	AT140253
C	7	407.4	17.4	936	10	BF126356
C	8	396.4	16.9	658	14	BY742680
C	9	394.4	16.9	417	8	AV695922
C	10	365	15.6	471	12	BM856449
C	11	360	15.4	360	9	AV660973
C	12	314.4	13.4	794	14	CN316999
C	13	309.8	13.2	393	9	AT597378
C	14	307.2	13.1	424	9	AA656720
C	15	297	12.3	336	6	TC93792
C	16	288.2	12.3	764	12	B1246567
C	17	286	12.2	722	10	BB667343
C	18	281.4	12.0	374	9	AA511669
C	19	281	12.0	460	9	AA239884
C	20	266.4	11.4	356	9	AA239884
C	21	266.4	11.4	356	9	AA239884
C	22	211	10.0	331	14	T86385
C	23	207	9.8	625	9	AA112016
C	24	206	8.8	502	9	BB869579
C	25	206	8.8	502	9	AA237916
C	26	203.6	8.7	706	29	AG094162
C	27	199.2	8.5	3623	11	AK004871
C	28	195	8.3	275	9	AT592875
C	29	188.4	8.1	606	14	CD502116
C	30	187	8.0	783	13	BX092855
C	31	179.4	7.7	516	9	AA237183
C	32	178.6	7.6	366	14	TC93842
C	33	172.4	7.4	2417	11	AK050938
C	34	161	6.9	1245	14	CD502117
C	35	147.4	6.3	14	CD502117	
C	36	147.2	6.3	356	14	CB813243
C	37	142.2	6.1	442	9	CB812866
C	38	141.6	6.1	442	29	AA217272
C	39	139.8	6.0	833	10	AG122753
C	40	135.4	5.8	610	12	BF650664
C	41	135.4	5.8	749	13	BM725726
C	42	125.8	5.4	527	9	BU740564
C	43	125.8	5.4	816	10	AT497110
C	44	125	5.3	865	10	BC299553
C	45	122	5.2	855	28	AZ051299
						B2637045
						OGCAN087C
						AT103358
						OX2f10.s
						AT140253
						qe2ia04.x
						BF126356
						6017693070
						BY742680
						AV695922
						AV695922
						AV695922
						BM856449
						K-EST014D
						AV660973
						AV660973
						CA316999
						UI-M-FW0
						AT597378
						vj29c06.y
						AA656720
						vp95e08.x
						TC93792
						ye05f01.s1
						B1246567
						602958477
						BB667343
						BB667343
						AA511669
						vj29c06.r
						AA239884
						mx81d01.r
						T86385
						yd77b08.r1
						AV695922
						TC93792
						ye05f01.s1

ALIGNMENTS

[illegible]

VERSION AV689089.1 GI:10290952
KEYWORDS EST.

SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE
AUTHORS
1 (bases 1 to 594)
Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.

TITLE

JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
MEDLINE
21625106

COMMENT

Contact: Zequang Han
Chinese National Human
Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; .

1726 GGGGCTGCTTGTGGATCTGGATTCCAGAAACATACAGAAATGCCCATTCCTTTTAA 1785


```

Db      477  GNGGTGCTTGTGATGCTGATTCCTCAGAAACATACAGAAATGCCATCTTTTAA 536

RESULT 3
AV720911/c
LOCUS   AV720911 477 bp mRNA linear EST 16-OCT-2000
DEFINITION AV720911 GLC Homo sapiens cDNA clone G1CETC06 5', mRNA sequence.
ACCESSION AV720911
VERSION   AV720911.1 GI:10818063
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 477)
AUTHORS Qian,B., Wu,T., Huang,Q., Kang,B., Gao,X., Xu,Z., Xiao,H.,
Xu,S., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L.,
X., Gu,Y., Chen,Z. and Han,Z.
XU,S., GU,M., TU,Y., JIA,J., FU,G., REN,S., ZHONG,M., LU,G., YANG
Homo sapiens cDNA GLC clones
JOURNAL Unpublished
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
Source
1. 477
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="G1CETC06"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="adult"
/lab_host="SOLR"
/clone_lib="GLC"
/notes="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 148 a 103 c 95 g 131 t
ORIGIN
1843 GTTTCAGGACGATGATTTCTGCTGGCAGCTCAAAATGTTCTGTGACAACATCAAT 1902
477 GTTTCAGGACGATGATTTCTGCTGGCAGCTCAAAATGTTCTGTGACAACATCAAT 418
1903 GTGTCCTTCACATCAAGAAATTCATTCAGAAAGAACTGCCAGGTGCACATCTAG 1962
417 GTGTGCTTCATCAAGAAATTCATTCAGAAAGAACTGCCAGGTGCACATCTAG 358
1963 ATTCACAAATGAACTTCGATTTGATTTATTCATTTATTCAGCTCTTGATCTAGGAT 2022
357 ATTCACAAATGAACTTCGATTTGATTTATTCATTTATTCAGCTCTTGATCTAGGAT 298
2023 AGTTTTCATCAAAATTAAGGATCAATCTATTAGAGAGTAAAGCCATGGCTGGAAA 2082
297 AGTTTTCATCAAAATTAAGGATCAATCTATTAGAGAGTAAAGCCATGGCTGGAAA 238
237 ATGGAAGTGAAGCTCCGACCTGTGATGATCTGTAACGCTCAAAATGAGAGTGCAT 178
2083 ATGGAAGTGAAGCTCCGACCTGTGATGATCTGTAACGCTCAAAATGAGAGTGCAT 2142
2143 GATTTTCCTTCGACAGACATCTCAAGTCTTTTAACTTAAGACATCCATTTGGCT 2202
177 GATTTTCCTTCGACAGACATCTCAAGTCTTTTAACTTAAGACATCCATTTGGCT 118
2203 CTTCGATTCAGACAGGCTTGAATGCAATGGAAGTGTATTAGTCCCTCTCTTACA 2262

Query Match 20.4%; Score 477; DB 9; Length 477;
Best Local Similarity 100.0%; Pred. No. 6,1e-112;
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      117  CTTCGATTCAGACAGGCTTGAATGCAATGGAAGTGTATTAGTCCCTCTCTTACA 58

RESULT 4
BB598373
LOCUS   BB598373 713 bp mRNA linear EST 26-OCT-2001
DEFINITION BB598373 RIKEN full-length enriched, adult male liver tumor Mus
musculus cDNA clone C730003G04 5', mRNA sequence.
ACCESSION BB598373
VERSION   BB598373.2 GI:16450340
KEYWORDS EST.
SOURCE   Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 713)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiromoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Kono,H., Kouda
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
Tagami,M., Tagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished
COMMENT On Dec 1, 2000 this sequence version replaced gi:11506974.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.c.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Matahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multichannel sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Location/Qualifiers
1. 713
/organism="Mus musculus"
/mol_type="mRNA"

```


QY	1918	AGGAATTCATTGAGAAAACTGCCAGGTGCACATCTAGATTACAAATGACTT	1977
Db	418	AGGAATTCATTGAGAAAACTGCCAGGTGCACATCTAGATTACAAATGACTT	359
QY	1978	TCGATTTGGTATTCATTTATTCACAGCTTTGTCACTCTAGAAATAGTTGTTTCAAAAT	2037
Db	358	TCGATTTGGTATTCATTTATTCACAGCTTTGTCACTCTAGAAATAGTTGTTTCAAAAT	299
QY	2038	AAGGATTCATTCATTAGCAGGTAGGAAAGCATGGCTGGGAAATGGAAGTGAAGCTG	2097
Db	298	AAGGATTCATTCATTAGCAGGTAGGAAAGCATGGCTGGGAAATGGAAGTGAAGCTG	239
QY	2098	CCGACTGTGCATGACTGCTGTGMAAGCTGTGAAATAGAGTGCATGATTTCTTTCTTGA	2157
Db	238	CCGACTGTGCATGACTGCTGTGMAAGCTGTGAAATAGAGTGCATGATTTCTTTCTTGA	179
QY	2158	CAGGACATCTCAAGCTTTTAAACCATTAAGACTCCATTTGCGCTTGTGATCCAAAGCAG	2217
Db	178	CAGGACATCTCAAGCTTTTAAACCATTAAGACTCCATTTGCGCTTGTGATCCAAAGCAG	119
QY	2218	GCGTTGAATGCAATGAAGTGGTTATATAGTCCCTTGCCTTACAACTTGACAGGACATGT	2277
Db	118	GCGTTGAATGCAATGAAGTGGTTATATAGTCCCTTGCCTTACAACTTGACAGGACATGT	59
QY	2278	GGTATTTTGGAAATGTGACTGACGCGGACCCCAAGATGTAATATTCATAAACTT	2335
Db	58	GGTATTTTGGAAATGTGACTGACGCGGACCCCAAGATGTAATATTCATAAACTT	1
RESULT 7	BF162656	936 bp	EST 30-OCT-2000
LOCUS	601769307F1 NCI_CGAP_Lu29	Mus musculus cdna clone IMAGE:398877 5'	
DEFINITION	nrna sequence.		
ACCESSION	BF162656		
VERSION	BF162656.1	GI:11042879	
KEYWORDS	EST.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 936)		
JOURNAL	NIH-MDC http://mgc.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished		
	Contact: Robert Strausberg, Ph.D.		

```

FEATURES
source
    Tissue Procurement: Gilbert Smith, Ph.D
    CDNA Library Preparation: Life Technologies, Inc.
    CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLt)
    DNA Sequencing by: Incyte Genomics, Inc.
    Clone distribution: MGC clone distribution information can be
    found through the I.M.A.G.E. Consortium/LNLt at:
    http://image.llnl.gov
    Plate: L1AW9197 row: m column: 02
    High quality sequence step: 686.
    Location/Qualifiers
        1. 936
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="CZECH 11"
            /db_xref="taxon:10090"
            /clone="IMAGE:398877"
            /tissue_type="spontaneous tumor, metastatic to mammary.
            Stem cell origin."
            /lab_host="DH10B"
            /clone_lib="NCI CGAP Lu29"
            /note="Organ: lung; Vector: pCMV-SPORT6, Site_1: SalI;
            Site_2: NotI; Cloned unidirectionally. Primer: oligo dT
            Library constructed by Life Technologies. Investigator
            providing samples: Gilbert Smith, NIH"

```


Query Match	16.9%	Score	396.4	DB	14	Length	658
Best Local Similarity	78.2%	Pred. No.	4.9e-91				
Matches	513	Conservative	0	Mismatches	138	Indels	5
						Gaps	3

QY	90	TTTGCCTGCCTTTGGCCATGGGTGACCTCTCATCTTTTACACCCCGGAGGGTCCATGGGTC	149
Db	2	TTTCTTCTTCTCTGATGGCTGGGTGAGTGCCCTTCTGTAGTCCAGAGGAGCCACAAGGC	61
QY	150	TTCCAATTAACAGAGGCTCCAGAGCTCCTTGGAGGGGGCTTCGCAACCCGCGCGAGC	209
Db	62	CTCACATCAACAGAGGGCTCTGTGAGCTCCCTGGAGCAAGGTTGGCTCACGGGACACAGAG	121
QY	210	CT---CACAGCCTGGGCACTCCTCATGCTCCTTACAGAGGTACGACCGCGGTGAGCCCT	266
Db	122	CTGGGCAACAGTTTAGTGTCCTCGATGTGTCTCTACAGCTCAGCAACCGTGTGGGGCCTT	181
QY	267	GGTGGGCAATCACATCTTCCCGGACAGCAGTGGACCGAGATCCTCAAGATGTCTCT	326
Db	182	GGTGGAACTCAATCAATGCCAGAGAGTGGACAGGCAAAATCCTCAAGATGTCTCT	241
QY	327	TGTACGTGGAGAGGGGAGATCATGTGCAATCTTAAGAAAGTCAAGAGCTCCGGGAAACCA	386
Db	242	TGTATCATTCAGAGAGTGGCCAGATTATGTGCACTTTAAGGAGGTCAAGGCTCAGGGAAACCA	301
QY	387	CGCTGCTGGAAGCCCATGTCCGGGAGGCTGGGGGCGCGGGGAGACCTTCTTGGGGAGGTGT	446
Db	302	CGGTGCTGGAGAGCCATCTCCGGGAGGCTGGCGGCGCACATGGGACCCCTGGAAGGGAGGTGT	361
QY	447	ATGTGAACGGGCGGGGGCTGGGGGAGGAGAGAGTTCAGAGACTGCTTCTCTAGTCTCTGC	506
Db	362	TTTGTGAATGGCTGTGAGACTGTGGGAGGAGACCAAGTTTCAAAGACTGTCTTCTCTACGCTCTGC	421
QY	507	AGAGCGACACCCCTGTCTGAGCAGCTCTACCGTGTGCGAGAGCGCTCACTATCACCCGCGTGC	566
Db	422	AGAGCGACGTTTTTTCTGAGCAGCCTCACTGTGGCGAGAGCTTGCGATACACAGCATGC	481
QY	567	TGGCCATCCGCGCGCGGATATCCCGGCTCTTTCCAGAAAGAAAGTGGAGGCGGCTCATYGCAG	626
Db	482	TGGCCCTCTCCGCGAGCTCTCGCGGACTTCTACAAACAAGAGGTGAGGCGAGTCAATGACAG	541
QY	627	AGCTAGATCTGACCATGTGGAGAGCGGACTGATTTGGCACTATACAGTTGGGGGCAATTT	686
Db	542	AGCTGAGGCTGAGCCAGCAGTGGCGGACCANNATGATTTGGACAGCTATTAATTTTGGGGG-ATNT	600
QY	687	CCACGGGTGAGCGGCGCGGGTCTCCATGCGACGCCAGTGTCTCAGAGATCTTAAG	742
Db	601	CCAGTGGCGAGCGGCGCGGAGT--TCATCTCCAGCGCCAACTTCTTAGAGACCCCAAGG	655

RESULT		9		
LOCUS	AV695922			
DEFINITION	AV695922 GKC Homo sapiens cDNA clone GKCDWE04 5'	417 bp	mRNA	linear EST 16-JAN-2002
VERSION	AV695922			
KEYWORDS	AV695922.1 GI:10297785			
SOURCE	EST.			
ORGANISM	Homo sapiens (human)			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 417)			
AUTHORS	Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X., Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,Y., Hu,W., Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X., Hu,G., Gu,Y., Chen,Z., and Han,Z.			
TITLE	Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	98 (26),	15069-15094	(2001)
MEDLINE	PUBMED	21625106		
COMMENT	Contact: Zeguang Han Chinese National Human Genome Center at Shanghai 331 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 1203, P. R. China Tel.: 86-21-50801919(ex.45) Fax: 86-21-50801922 Email: hanzg@chgc.sh.cn This clone is available at CHGC in Shanghai. Location/Qualifiers			
FEATURES				

```

source
1. .417
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="GKCDME04"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GKC"
/notes="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      70 a      114 c      109 g      124 t
ORIGIN

```

Query Match	16.9%	Score 394.4	DB 9	Length 417
Best Local Similarity	97.9%	Pred. No. 1.3e-90		
Matches 411	Conservative 0	Mismatches 6	Indels 3	Gaps 1
1246 GAGAAATAGCTGCAGTATTAACGGGTCTCCCTTGAGATTCGATCATGATGGTTTGTCTCT				1305
1 GAGAAATAGCTGCAGTATTAACGGGTCTCCCTTGAGATTCGATCATGATGGTTTGTCTCT				60
1306 CCTTTCTTCGTTCTGCGGGTCCGAGCAATGTGCTAAAGGGTGTATCCAGACCGGT				1365
61 CCTTTCTTCGTTCTGCGGGTCCGAGCAATGTGCT--GGTGTATCCAGACCGGT				117
1366 AGGTCTCCTTTACCAATTTGTGGGGCCACCCTGACACAGGCATGCTGAACGTGTGAA				1425
118 AGGTCTCCTTTACCAATTTGTGGGGCCACCCTGACACAGGCATGCTGAACGTGTGAA				177
1426 TCTGTTTCCCGTCTGCGAGCTGTACGACACAGAGAGTACGAGCGGCTCTACACAGA				1485
178 TCTGTTTCCCGTCTGCGAGCTGTACGACACAGAGAGTACGAGCGGCTCTACACAGA				237
1486 GTGGCAGATGATGCTGGGCTATGACACTGACACTCTCCCTTCACGCTTTGGCCACCAT				1545
238 GTGGCAGATGATGCTGGGCTATGACACTGACACTCTCCCTTCACGCTTTGGCCACCAT				297
1546 GATTTTTCAGCATGTGTGCTACTGAGACGTGGGCTTAATCTGTAGAGTTGCCCATTTGG				1605
238 GATTTTTCAGCATGTGTGCTACTGAGACGTGGGCTTAATCTGTAGAGTTGCCCATTTGG				357

Oy 1606 ATATTTTCTGCTGCTCTTGGCCCCCCTTAATGCTAATTTCTAACTCTTGCT 1665
 Db 358 ATATTTTCTGCTGCTCTTGGCCCCCCTTAATGCTAATTTCTCTTGCTCT 417
 RESULT 10
 BM856449
 LOCUS BM856449 471 bp mRNA linear EST 06-MAR-2002
 DEFINITION K-EST0140406 S14K402 Homo sapiens cDNA clone S14K402-48-E04 5',
 mRNA sequence.
 ACCESSION BM856449
 VERSION BM856449.1 GI:19212848
 SOURCE EST.
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 471)
 Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.D., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
 Kim,Y.S.
 21C Frontier Korean EST Project 2001
 JOURNAL Unpublished
 COMMENT Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 48 row: E column: 04
 High quality sequence stop: 471.
 Location/Qualifiers
 1..471
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S14K402-48-E04"
 /cell_line="RK402"
 /lab_host="TOP10P"
 /note="Organ: Stomach; Vector: pTZ189P1; Site_1: EcoRI;
 Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then deapped
 with tabacco acid pyrophosphatase (TAP). The deapped
 intact mRNA was ligated with DNA-RNA linker including EcoR
 I site by treatment of T4 RNA ligase and the first strand
 cDNA was synthesized from oligo dT-selected mRNA by
 priming with dT-tailed vector. The dT-tailed vector was
 adjusted to have about 60nt. The cDNA vector was
 circularized with E. coli DNA ligase after digestion of
 EcoRI which site is also included in vector. An RNA strand
 converted to a DNA strand by Okayama-Berg method. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli TOP10P by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library."
 BASE COUNT 138 a 79 c 101 g 153 t
 ORIGIN
 Query Match 15.6%; Score 365; DB 12; Length 471;
 Best Local Similarity 100.0%; Pred. No. 5.5e-83;
 Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1976 TTCTGATTTGTTATTCAGCTCTTCTCATCCAGAAATAGTGTTCAAA 2035
 Db 1 TTCTGATTTGTTATTCAGCTCTTCTCATCCAGAAATAGTGTTCAAA 60
 Oy 2036 ATAAGGATCATCTCATTAGCAGTACTGAAGCCATGCGTGGAAATGGAAGTGAAC 2095
 Db 61 ATAAGGATCATCTCATTAGCAGTACTGAAGCCATGCGTGGAAATGGAAGTGAAC 120

Oy 2096 TCCGACATGTCAGTACTGCTCTGACCTGTGAATGAGAGTGCATATTTCTTCT 2155
 Db 121 TGCCGACATGTCAGTACTGCTCTGACCTGTGAATGAGAGTGCATATTTCTTCT 180
 Oy 2156 GACAGGATCTCAAGCTCTTTAACCATTAAGACTCCATTGTGCTCTTGATCAAC 2215
 Db 181 GACAGGATCTCAAGCTCTTTAACCATTAAGACTCCATTGTGCTCTTGATCAAC 240
 Oy 2216 AGCGCTGATGCAATGAGAGTGTTPAGCCCTGCTCTTCAACTTCAGGACAT 2275
 Db 241 AGCGCTGATGCAATGAGAGTGTTPAGCCCTGCTCTTCAACTTCAGGACAT 300
 Oy 2276 GTGCTTATTTGAAATTTGACTGACGCGACCAAGATGTAATATTCATTAACCT 2335
 Db 301 GTGCTTATTTGAAATTTGACTGACGCGACCAAGATGTAATATTCATTAACCT 360
 Oy 2336 ATGGG 2340
 Db 361 ATGGG 365
 RESULT 11
 AV660973
 LOCUS AV660973 360 bp mRNA linear EST 16-JAN-2002
 DEFINITION AV660973 GLC Homo sapiens cDNA clone GLCNC08 3', mRNA sequence.
 ACCESSION AV660973
 VERSION AV660973.1 GI:9881987
 SOURCE EST.
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 360)
 Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
 Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
 Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
 Hu,G., Gu,J., Chen,Z. and Han,Z.
 Insight into hepatocellular carcinogenesis at transcriptome level
 by comparing gene expression profiles of hepatocellular carcinoma
 with those of corresponding noncancerous liver
 Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
 JOURNAL Chinese National Human Genome Center at Shanghai
 MEDLINE 21625106
 PUBMED 11752456
 COMMENT Contact: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.
 Location/Qualifiers
 1..360
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="GLCNC08"
 /tissue_type="corresponding non cancerous liver tissue"
 /dev_stage="Adult"
 /lab_host="SOLR"
 /clone_id="GLC"
 /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
 XhoI"
 BASE COUNT 90 a 77 c 70 g 123 t
 ORIGIN
 Query Match 15.4%; Score 360; DB 9; Length 360;
 Best Local Similarity 100.0%; Pred. No. 9.5e-82;
 Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1577 GCCTTACATCTTGAAGTTGCCGATTTGGAATTTTCTGCTCTTGCCGCCCCAC 1636
 Db 1 GCCTTACATCTTGAAGTTGCCGATTTGGAATTTTCTGCTCTTGCCGCCCCAC 60

QY 1637 TTAATTGGTAATTTCTTAACCTTGTGCTACTTGTGATCGTCCAAATCCAAATATAGTC 1696
 |||||||
 Db 61 TTAATTGGTAATTTCTTAACCTTGTGCTACTTGTGATCGTCCAAATCCAAATATAGTC 120
 |||||||
 QY 1697 AACAGGTAGTGGCTCTGCTGTCCATGCGGGGGCTCTGTGGATCGATTCCTCAGA 1756
 |||||||
 Db 121 AACAGGTAGTGGCTCTGCTGTCCATGCGGGGGCTCTGTGGATCGATTCCTCAGA 180
 |||||||
 QY 1757 AACATACAGAAATGCCATTCCTTTAAATCATCAGTATTTTACATTCCAAATAT 1816
 |||||||
 Db 181 AACATACAGAAATGCCATTCCTTTAAATCATCAGTATTTTACATTCCAAATAT 240
 |||||||
 QY 1817 TCCAGTACAGATTCCTGTACTCATGATGCTTACGAGCTGAATTTCACTGTGGACCTCA 1876
 |||||||
 Db 241 TCCAGTACAGATTCCTGTACTCATGATGCTTACGAGCTGAATTTCACTGTGGACCTCA 300
 |||||||
 QY 1877 AATGTTTCTGTACACATTCATTCATGTGTGCTTCACTCAAGAAATTCATTCATTCAG 1936
 |||||||
 Db 301 AATGTTTCTGTACACATTCATTCATGTGTGCTTCACTCAAGAAATTCATTCATTCAG 360
 |||||||

RESULT 12
 CA316999 794 bp mRNA linear EST 26-NOV-2002

LOCUS CA316999
 DEFINITION UI-M-FW0-cbm-a-08-0-UI.r1 NIH_BMAP_FW0 Mus musculus cDNA clone
 IMAGE: 6811377 5', mRNA sequence.

ACCESSION CA316999
 VERSION CA316999.1 GI:24535123
 KEYWORDS EST.

SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE NIH-MGC http://mgc.ncl.nih.gov/.
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-rrmail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LIMB at:
 http://image.liml.gov

FEATURES
 source
 Seq primer: pyx-5.
 Location/Qualifiers
 1..794

/organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE: 6811377"
 /tissue_type="whole brain"
 /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_id="NIH_BMAP_FW0"
 /note="Organ: Brain; Vector: pyx-Asc; Site: 1; Ecor I;
 Site 2: Not I; The library was constructed according
 to Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with Ecor I adaptor, digested with NotI and then cloned
 directionally into pyx-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is AGCAGACAG. This library was created for the University
 of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the

BASE COUNT 156 a 239 c 220 g 176 t 3 others
 ORIGIN
 Query Match 13.4%; Score 314.4; DB 14; Length 794;
 Best Local Similarity 71.8%; Pred. No. 8.5e-70;
 Matches 481; Conservative 0; Mismatches 103; Indels 86; Gaps 2;

370 AGGCTCCGGAAACACACCTGCTGAGACCCATGTCGGGAGAGCTGGGGCGGGGAGC 429
 |||||
 Db 87 AGGCTAGGAAAGACACACCTGCTGAGACCCATGTCGGGAGAGCTGGGGCGGGGAGC 146
 |||||
 QY 430 CTTCCTGGGGAGGTGTATGTAAAGCGCGGGCGCTGCCCGGAGACAGTCCAGAGCTG 489
 |||||
 Db 147 CCTGGAGGGGAGGTGTATGTAAAGCTGTGCAATGCTGTGCGAGCTGCCAGGAGACAGTCCAGAGCTG 206
 |||||
 QY 490 CTTCCTGCTGCG----- 500
 |||||
 Db 207 CTTCCTGCTGCTGCTGCGAGGTGGCGGTGCTCCCTGCGCCCTGCGCGGCTGCGGCC 266
 |||||
 QY 501 -----TCTTGAGAGCGACACCCCTGCTGA 524
 |||||
 Db 267 CTAGCCCGCGGATTTGACGACACCTGATGTCCTTCTCTGAGAGCGAGCTTTTCTGA 326
 |||||
 QY 525 GCAGCCTCACCGCTGGCGCGAGAGCTGACATACACCGCGCTGCGCCATCCCGCGGCA 584
 |||||
 Db 327 GCAGCCTCACCTGCGCGAGAGCTGCGATACACAGCGATGCTGCGCCCTGCGCGAGCT 386
 |||||
 QY 585 ATCCCGGCTCCTTCAGAAAGAGTGGAGCGCGTATGCGAGAGCTGATGAGCCATG 644
 |||||
 Db 387 CCGCGAGCTCTTACAAAGAGTAAAGCGAGTATGACAGAGCTGAGAGCCATGAGCCACG 446
 |||||
 QY 645 TGGCAGACGAGCTGATGAGCAACTACAGCTGGGGGGGATTTCCACAGGTTGAGCGCGCG 704
 |||||
 Db 447 TGGCGGCAATATGATTTGGAGCTTAAATTTGGGGAAATTTCCAGTGGCGCGCGCGCG 506
 |||||
 QY 705 GGGTCTCATGCGACCGACCTGCTCCAGAGATCTTAAGCTATGCTGTTGATGAGCCAA 764
 |||||
 Db 507 GAGTTCCATGCGACCGACCGACCTGCTCCAGAGATCTTAAGCTATGCTGTTGATGAGCCAA 566
 |||||
 QY 765 CCACAGGCGCTGAGCTGAGTACTGCTATGAGATGCTGCTCTGCTGAGAGCTGCTC 824
 |||||
 Db 567 CCACAGGCGCTGAGCTGAGTACTGCAAAATTAATTTCTTGTGCTGAGCTGCTC 626
 |||||
 QY 825 GCAGAGACGAAATGCTGCTGCAACATTCACACCGCGCTTGTAGGTTTTCAGCTCT 884
 |||||
 Db 627 GCAGAGACGAAATGCTGCTGCAACATTCACACCGCGCTTGTAGGTTTTCAGCTCT 686
 |||||
 QY 885 TTGACAAATTCGATCTGAGCTGAGCTGAGAGCTGATTTTGTGAGAGCGAGAA 944
 |||||
 Db 687 TCGACANAATTCGATCTGAGCTGAGAGCTGAGAGCTGATTTTGTGAGAGCGAGAA 745
 |||||
 QY 945 TCGTGTGATT 954
 |||||
 Db 746 TCGTGTGATT 755
 |||||

RESULT 13
 A1597378
 LOCUS A1597378 393 bp mRNA linear EST 21-APR-1999
 DEFINITION vj29c06.y1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA
 clone IMAGE:930442 5', mRNA sequence.
 ACCESSION A1597378
 VERSION A1597378.1 GI:4606426
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 REFERENCE
 AUTHORS Matra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,

Query Match 12.7%; Score 297; DB 14; Length 336;
Best Local Similarity 96.9%; Pred. No. 1.8e-65;
Matches 312; Conservative 0; Mismatches 8; Indels 2; Gaps 1;

Search completed: July 27, 2003, 08:29:24
Job time : 4356.59 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 09:11:47 ; Search time 18-Seconds

(without alignments)
1064.129 Million cell updates/sec

Title: US-09-989-981a-6

Perfect score: 3336

Sequence: 1 MGDLSLTPGSGMGLQVNRG.....PALVILGIVEKINDHLIR 651

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents.AA.*
2: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/6C.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	682.5	20.5	655	4 US-09-245-808-1	Sequence 1, Appl1
2	244	7.3	1280	2 US-08-134-001C-3731	Sequence 3731, Ap
3	244	7.3	1280	2 US-08-752-447-2	Sequence 2, Appl1
4	244	7.3	1280	4 US-09-316-167-2	Sequence 2, Appl1
5	240	7.2	1279	2 US-08-784-649A-2	Sequence 2, Appl1
6	240	7.2	1280	2 US-08-583-276-19	Sequence 19, Appl1
7	240	7.2	1280	6 5206352-4	Patent No. 5206352
8	239.5	7.2	1684	3 US-08-665-259-25	Sequence 25, Appl1
9	239.5	7.2	1684	3 US-08-762-500-25	Sequence 25, Appl1
10	239.5	7.2	1704	3 US-08-762-500-75	Sequence 75, Appl1
11	230.5	6.9	242	4 US-09-134-001C-3832	Sequence 3832, Ap
12	229	6.9	460	4 US-09-134-001C-3369	Sequence 3369, Ap
13	229	6.9	1307	1 US-08-395-246C-2	Sequence 2, Appl1
14	220.5	6.6	1273	4 US-09-134-001C-3641	Sequence 3641, Ap
15	220	6.6	1457	3 US-08-665-259-27	Sequence 27, Appl1
16	220	6.6	1457	3 US-08-762-500-27	Sequence 27, Appl1
17	218	6.6	1275	4 US-09-450-105-2	Sequence 2, Appl1
18	218	6.6	1275	4 US-09-450-105-2	Sequence 2, Appl1
19	213	6.4	228	4 US-09-134-001C-4373	Sequence 4373, Ap
20	211	6.3	270	4 US-09-134-001C-5561	Sequence 5561, Ap
21	210	6.3	382	4 US-09-134-001C-3439	Sequence 3439, Ap
22	209.5	6.3	244	4 US-08-919-573-4	Sequence 4, Appl1
23	209.5	6.3	244	4 US-08-919-573-4	Sequence 4, Appl1
24	209.5	6.3	341	4 US-09-134-001C-4042	Sequence 4042, Ap
25	208.5	6.3	202	4 US-08-858-207A-332	Sequence 332, Ap
26	208.5	6.3	229	4 US-09-134-001C-4113	Sequence 4113, Ap
27	207.5	6.2	246	4 US-09-627-376-15	Sequence 15, Appl1

28	206.5	6.2	233	4 US-08-993-825-2	Sequence 2, Appl1
29	206.5	6.2	1476	4 US-09-256-703-2	Sequence 2, Appl1
30	206.5	6.2	1480	1 US-07-637-621-2	Sequence 2, Appl1
31	206.5	6.2	1480	2 US-08-951-912-2	Sequence 2, Appl1
32	206.5	6.2	1480	4 US-08-681-838A-2	Sequence 2, Appl1
33	206.5	6.2	1480	4 US-08-681-838A-3	Sequence 3, Appl1
34	206.5	6.2	1480	4 US-09-174-077-2	Sequence 2, Appl1
35	206.5	6.2	1480	6 5240846-5	Patent No. 5240846
36	206	6.2	406	4 US-09-134-001C-3570	Sequence 3570, Ap
37	206	6.2	1479	2 US-08-951-912-4	Sequence 4, Appl1
38	206	6.2	1479	4 US-09-174-077-4	Sequence 4, Appl1
39	205.5	6.2	1480	4 US-09-425-453A-6	Sequence 6, Appl1
40	205.5	6.2	1480	4 US-09-425-453A-12	Sequence 12, Appl1
41	205.5	6.2	1480	4 US-09-425-453A-20	Sequence 20, Appl1
42	205	6.2	225	4 US-09-134-001C-3215	Sequence 3215, Ap
43	204.5	6.1	1308	2 US-08-996-644-2	Sequence 2, Appl1
44	204.5	6.1	1308	3 US-09-352-552-2	Sequence 2, Appl1
45	204.5	6.1	1480	1 US-08-136-742A-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1	
US-09-245-808-1	
Sequence 1, Application US/09245808	
Patent No. 6313277	
GENERAL INFORMATION:	
APPLICANT: Doyle, L. Austin	
APPLICANT: Abruzzo, Lynne V.	
APPLICANT: Ross, Douglas D.	
TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which	
TITLE OF INVENTION: encodes it	
FILE REFERENCE: Ross UMD conversion	
CURRENT APPLICATION NUMBER: US/09/245, 808	
CURRENT FILING DATE: 1999-02-05	
EARLIER APPLICATION NUMBER: 60/073763	
EARLIER FILING DATE: 1998-02-05	
NUMBER OF SEQ ID NOS: 7	
SOFTWARE: PatentIn Ver. 2.0	
SEQ ID NO: 1	
LENGTH: 655	
TYPE: PRT	
ORGANISM: Human MCF-7/AdrVP cells	
US-09-245-808-1	
Query Match	20.5%; Score 682.5; DB 4; Length 655;
Best Local Similarity	29.2%; Pred. No. 1.9e-64;
Matches 182; Conservative 138; Mismatches 249; Indels 55; Gaps 18;	
21 SSSLSGAPATAP---EPHSLGILHASYSVSHRRPMDITSGROQWTROIKDVSLYE 77	
13 SGGTNGFPAATASNDLKAFTGAVLSPHNICRYNKLKSGFLPCRKPVEKEILSNINIMK 72	
78 SQGIMCLSSGSGGKTTILDAMSGRLGACFTFGEVYVNGRALRREQFDQCFSVYLSQT 137	
73 PG-LNMLIGPTGGKGLDVLVLAARKDPG-LSGDVLING-APRPANFKCNSSGVVQDDV 129	
138 LLSLIVREFLHTALLAIRGNPG--SFQKVAVMAELSLSHVADRLTINYSIGISTG 196	
130 VMGTLVIREKHSALIRLATYTTNHEKNRINRVIOELGLDVAASKVTGPIRIGSG 189	
197 ERRRVIAQLADPRVMLDEFTTGLDCTANQIVLVLELARRNRIVLTHOPRSEL 256	
190 ERKRTSIGMLTDPISLFLDEFTTGLDSTANAVALLLRMSKGGRTTIFSIHQPRYSI 249	
257 FOLFALIALSLGELIFCGTPAEMLDFFNDGCGPCPEHNSPFEDYDLTSVDTQ---SK 312	
250 FKLFSLTILTAGRLMFGPAQEAALGYFSAAGYHCAVYNNPDEFDIINGSTAVANLR 309	
313 ERE-----IETSKR-----VOMIESAYKKSALCHKT-----LKNTERKKHLKTLPMVPP 356	
310 EEDFKATELIEPSKQDKPLIEKLALEYVNSSEYFKETKALHQLSGEKKKKITVFEIERY 369	

QY 357 KTKDSFVSKLGVLKRVTRVLNKLAVITRLONLIMGLFL--LFFVLKRVSNVIMG 414
 DB 370 TT-----SPCHOLRWKSKRSFKMLGNPOASIAQIYTVVLGVLGAIYGLKNDST---- 421
 QY 415 AIDRVGLLYOFVGAAPYTGMLNAVLPVLKAVSQESODGLYKQWMLAYAL-HVLP 473
 DB 422 GIONRAGVLF-PLTTNOCSSVSAVELFVEKRLFHEXISGYRVSSEYFLKRLSDLP 480
 QY 474 FSVVAMIFSSVCYWTGLHPEVAREGYSAALLAPHLIGFELTVLLGIVONPNVNSV 533
 DB 481 KMLPFIITFCIVFYLGLKPKADAFVMMFTLM---MAYASASSALAIAGQSVSYVA 537
 QY 534 VALLSLAGV--LVSGFLNIQEMPIPKIISYFFQKCYSEILVYNEFYGLNFTGSSN 591
 DB 538 TLMITCFVFMIFSGLLVNLTTIASWLSMLQFSIPRYGFTALQHNERIEGONFCPG--- 594
 QY 592 VSVTPNPKCAFTGQIOFIEKTCPG 615
 DB 595 LNAVGNPCNVA-----FCTG 610

RESULT 2

US-09-134-001C-3731
 Sequence 3731, Application US/09134001C
 Patent No. 6380370
 GENERAL INFORMATION:
 APPLICANT: Lynn Doucette-Stamm et al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: GTC-007
 CURRENT APPLICATION NUMBER: US/09/134,001C
 CURRENT FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 60/064,964
 PRIOR FILING DATE: 1997-11-08
 PRIOR APPLICATION NUMBER: US 60/055,779
 PRIOR FILING DATE: 1997-08-14
 NUMBER OF SEQ ID NOS: 5674
 SEQ ID NO 3731
 LENGTH: 248
 TYPE: PRP
 ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-3731

Query Match 7.3%; Score 244; DB 4; Length 248;
 Best Local Similarity 29.3%; Pred. No. 8.4e-18;
 Matches 70; Conservative 53; Mismatches 92; Indels 24; Gaps 9;
 QY 67 QILKDVSLVESGQIMCLIGSSGSKTLLDAMSG-RIGRAGTFLGEVYNGRALRRE-- 123
 DB 20 EVIKIDIKINGOEYVTLIGRSGSKTLLIMINALIETEGT---YVNGMTYNTDK 75
 QY 124 ---OFODCFSYLQSDTLLSLVRETLHTALLAIRGNPQSFQKVEVAMLSLH 179
 DB 76 KSOIKVROOSGMFQNTYLFPHKSALENN-MGLTYVKMKKATANEAMMLAKVGLVH 134
 QY 180 VADRLIGNYSLGISTGERRRVSIAAQLDOPKVALFDEPTTGDCMTANOIVLVLETA 239
 DB 135 VKDQ--RPHALSG--GGQORVATARALANMPKXMLEPESALDPDELVYNLAKYKELA 189
 QY 240 RNRKIVLTITQPSSELPQFDKTAIISFGELIFCGTIPAEMLDFPNDGCPPEHSNPF 298
 DB 190 DEGMTVIVTHEMRPAK-EVSNQIAFIHEGVIAEGTPE--DIFN--HPKTEELQRF 241

RESULT 3

US-08-752-447-2
 Sequence 2, Application US/08752447
 Patent No. 5994088
 GENERAL INFORMATION:
 APPLICANT: Mechtner, Eugene
 APPLICANT: Roninson, Igor B
 TITLE OF INVENTION: Methods and Reagents for Preparing and

TITLE OF INVENTION: Using Immunological Agents Specific for F-glycoprotein
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: McDonnell Boenhen Hubbert & Berghoff Ltd.
 STREET: 300 South Wacker Drive, Seventh floor
 CITY: Chicago
 STATE: Illinois
 COUNTRY: USA
 ZIP: 60606
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/752,447
 FILING DATE: 15-NOV-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: No. 5994088nan, Kevin E
 REGISTRATION NUMBER: 35,303
 REFERENCE/DOCKET NUMBER: 95,1121
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 312-913-0001
 TELEFAX: 312-913-9808
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1280 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-752-447-2

Query Match 7.3%; Score 244; DB 2; Length 1280;
 Best Local Similarity 20.7%; Pred. No. 1.5e-16;
 Matches 153; Conservative 106; Mismatches 229; Indels 250; Gaps 32;
 QY 41 LHASYSHRYRPMMDISRCQWTRQLKDVSLVESGQIMCLIGSSGSKTLLDAMS 100
 DB 397 VHFSTPSKREYK-----ILKGIINTLKVGSGQVVALVNGSGCKSTVQIMQ 441
 QY 101 GRLGAGTFLGEVYNGRALR--REOFODCFSYLQSDTLLSLVRETLHTYLLAIR 158
 DB 442 RL-YDPTEGMVSVDGODITINRFLREILIGVSGQPVLFATIANIR----- 490
 QY 159 GNPQSFQKVAVNAE-----LSLHVADRLIGNYSLGISTGERRRVSIAAQLDOP 211
 DB 491 GREVTMDIEKAVKAEANAYDFIMKLPKFDLVGERS-AQLSGGQORAIARALVNP 549
 QY 212 KVMLEDEPTTGDCMTANOIVLVLELARNRIVVLTIHORSELPQFDKTAIISFEL 271
 DB 550 KILLDEPMSALD-TESAIVVQVALDKARKGRRTIVIAN--RFATVRADVIAGFDGVI 606
 QY 272 IFCGTPAEMDFPNDGCPPEHSNPFVMDLTSVDTSQKEREIETSK----- 320
 DB 607 VEKGNHDELM-----KEGCIYRKLTVMOTAGNEVELENAADSKSEIDALE 652
 QY 321 -----RVQMISAYKKSALCHTKLNIERMKLTLPPVFF-----RTKDSFG 363
 DB 653 MSNDSRSLIRKSTRSVSGSOAHKRLSTKALD--ESIPVSEFRIKMLNITEMPY 710
 QY 364 -----VESK-LGVLLR-----RVTRN-----LYRNLAIVYTR 389
 DB 711 FVGVFCAILINGIOPAPALIFSKIIIGVFTIIDPETRKONSLSFLIALGIISFTF 770
 QY 390 LLONLING-----LFLFVLYRVSNN-----LKGAIQ 417
 DB 771 FLGFTFGKAGEILTKRLRVVFRSMLRQDSWHDKNRTGALTTRLANDAAGYKKAIG 830
 QY 418 DRV-----GLIYQFVGATPYTGMLNAVNLPPVL-----RAVSDOE 452
 DB 831 SRLAVITQMIANLGTGIISFIYGMOLTLILLAI--VPIAIAGVEMKMFAGALDKK 888

QY 453 SODGL-----YOKOMMLAVLHV-----LPEFSVATM 480
 DB 889 ELEGKATATEAIENTFRVSLTOEQKEFHMTAOSLOVYRNSLRKAHIFGTFSTQAM 948
 QY 481 IFSSV--CYWTLGLHPEVARFGYFSALLAPHLIGEF--LTLVLGIYONPNIVNSV--- 533
 DB 949 MYFSTAGCF-----RFG---AYLVAKHLMSEFEDVLLVFSAVVFGAMAVQVSSF 994
 QY 534 -----VALSTAGVLVSGFELRNIOEMPIFKIISYFFFOKYCSEILVYNEFYGLNFTCG 588
 DB 995 APDYAKAKISAHHIM-----IIEKTP-----IDYSTEGIMPNTEG--NVTFG 1038
 QY 589 SSNSVTNPMCAFTOGI 606
 DB 1039 EVFENYPTRPDPIDVLOGL 1056

RESULT 4
 US-09-316-167-2

Sequence 2, Application US/09316167
 Patent No. 6365357
 GENERAL INFORMATION:
 APPLICANT: Mechtner, Eugene
 APPLICANT: Roninson, Igor B
 TITLE OF INVENTION: Methods and Reagents for Preparing and
 TITLE OF INVENTION: Using Immunological Agents Specific for P-glycoprotein
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: McDonnell Boenchen Hulbert & Berghoff Ltd.
 STREET: 300 South Wacker Drive, Seventh Floor
 CITY: Chicago
 STATE: Illinois
 COUNTRY: USA
 ZIP: 60606
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/316,167
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/752,447
 FILING DATE: 15-NOV-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: No. 6365357nan, Kevin E
 REGISTRATION NUMBER: 35,303
 REFERENCE/DOCKET NUMBER: 95,1121
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 312-913-0001
 TELEFAX: 312-913-9808
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1280 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-316-167-2

Query Match 7.3%; Score 244; DB 4; Length 1280;
 Best Local Similarity 20.7%; Pred. No. 1.5e-16;
 Matches 153; Conservative 106; Mismatches 229; Indels 250; Gaps 32;

QY 41 LAHSYSHRVPRMMDITSCRGQWROIKNDSLVYESGOIMCILLGSSGSKTLLDAMS 100
 DB 397 VHSYSRKEVK-----ILKGLNLTQVSGQTVALVNGSGCKSTVQLMQ 441
 QY 101 GRGGRAGTFLGEYVNGRRLR--REPODCCFTYLDSDTLSSLVRETLHYALLAIR 158
 DB 442 -RL--YDPTGEMVSDGQDIRTNVRFLEITIGVSOEPLVFTATTAENIRY----- 490

QY 159 GNPSPORKVEAVMAE-----LSLSHADLLIGNYSIGISTGERRRVSIAAQLDDP 211
 DB 491 GRENVTMDIEKAVEANAYDYMILPHFEDTLVBERG--AOLSGGOKORAIARLVRNP 549
 QY 212 KVMLEDEPTTGIDCMTANOIVLVLELARNRIVTLTIHQPSSELPFOLDKTAISFGL 271
 DB 550 KILLDEANSALD--TESAEVAVQVALDKARKGTTIVIAH--REAVYRNDVYIAGDDGYI 606
 QY 272 ITCGPAEMKDFPNOCIPCEHSPDFEDYMDLTSVQSKREIETSK----- 320
 DB 607 VEKGNHDELM-----REKGIYFLVMTQFAGNEVELENADESKSEIDALE 652
 QY 321 -----RVOMIESAYKKSACHTKTKIEMKHKITPMVPF-----KTKDSFG 363
 DB 653 MSSNDRSSLIRKSTRSVRSQAOHRRLSTKEALD--ESTIPVSFMRIMKLNLTMPY 710
 QY 364 -----VFSK-LGVLLR-----RVEN-----LVENKLAIVTR 389
 DB 711 FVYGVFCALINGLQPARALIESKILIGVTRIDDETRKONSLFSLFLALGITSFTIF 770
 QY 390 LLONLIMG-----LFLEFVLVRVSNV-----LKGAIQ 417
 DB 771 FLQGFTEPKAGEILTKRLRYWFRSMRLKODVSFHDPKNTGALTTRLANDAAYKGAIG 830
 QY 418 DRV-----GLIXQFGATPYTGMLNANVLPVL-----RAVSDOE 452
 DB 831 SRLAVITONIANLGTGIIISFYGMQLTLLAI--VPIIAGVEMKMPAGQALDKDK 888
 QY 453 SODGL-----YOKOMMLAVLHV-----LPEFSVATM 480
 DB 889 ELEGKATATEAIENTFRVSLTOEQKEFHMTAOSLOVYRNSLRKAHIFGTFSTQAM 948
 QY 481 IFSSV--CYWTLGLHPEVARFGYFSALLAPHLIGEF--LTLVLGIYONPNIVNSV--- 533
 DB 949 MYFSTAGCF-----RFG---AYLVAKHLMSEFEDVLLVFSAVVFGAMAVQVSSF 994
 QY 534 -----VALSTAGVLVSGFELRNIOEMPIFKIISYFFFOKYCSEILVYNEFYGLNFTCG 588
 DB 995 APDYAKAKISAHHIM-----IIEKTP-----IDYSTEGIMPNTEG--NVTFG 1038
 QY 589 SSNSVTNPMCAFTOGI 606
 DB 1039 EVFENYPTRPDPIDVLOGL 1056

RESULT 5
 US-08-784-649A-2

Sequence 2, Application US/08784649A
 Patent No. 5830697
 GENERAL INFORMATION:
 APPLICANT: Sikic, Branimir I
 APPLICANT: Chen, Gang
 TITLE OF INVENTION: P-GLYCOPROTEIN MUTANT RESISTANT TO
 TITLE OF INVENTION: CYCLOSPORIN MODULATION
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson
 STREET: 2200 Sand Hill Road
 CITY: Menlo Park
 STATE: CA
 COUNTRY: USA
 ZIP: 94025
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/784,649A
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Sherwood, Pamela J

```

: REGISTRATION NUMBER: Reg.No. 5830697 36,677
: REFERENCE/DOCKET NUMBER: 06037/007001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-322-5070
: TELEFAX: 415-854-0875
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1279 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-784-649A-2

```

```

Query Match 7.2%; Score 240; DB 2; Length 1279;
Best Local Similarity 20.7%; Pred. No. 4e-16;
Matches 152; Conservative 108; Mismatches 230; Indels 246; Gaps 32;

```

```

QY 41 LHASYSVSHRYRPMWDTSCROMTROIKDVSILYVESGOIMCTIGSSGSKTLLDAMS 100
DB 396 VHFSPSRKEVK-----ILKGLNLKVGSGQVALVGNCGCKSTTVQLMQ 440
QY 101 GLRLRAGTFLGEVYVNGRALR--REQFQDCSYVLQSDTLSSLTVRETLHTYLLAIR 158
DB 441 -RL--YDPTGMSVSDODIRITINVRFLREITIGVSOEPVLFATTIAENIRY----- 489
QY 159 GNPSPQKVEAVAAE-----LSLHVADRLIGNSLGISTGERRRYSIAQLODP 211
DB 490 GREVNTDEIEKAVKEANADYFMKLPHKFTLVGENG--AOLSGGOKRIATIALVNP 548
QY 212 KVMLEDEPTGLDCTANOIYVLLVLELARNRRIYVLTIHOPRSELPOLFDTIALISFGE 271
DB 549 KILLDEATSLD--TESAVVQVALDKARKGRITIVIAH--RLSTVRNADVIAGFDGVI 605
QY 272 IFCTPRMELDFNDGCPCEHNSNPDFYMDLTSVDTOSKERIEIYS-----KRVQ 323
DB 606 VEKGNHDELM-----KEKGIYFKLVTMOTAGNEVELENADESKSEIDALE 651
QY 324 MIESAYKKSATCHTKLNIERMKHL-----KTLPMVPE-----KTRDSPG-- 363
DB 652 MSSNDSSSLRKSTRSRVSGSOADRKSLTKREALDESIPVPSFMRIMKLNLTMPYEV 711
QY 364 -----VESK--LGVLRL-----RVTRN-----LVNRKLAVITRL 391
DB 712 VGVECAIINGGLAPAFALIFSKIGVFTRIDDPETKRONSLFSLFLALGIIISFTFF 771
QY 392 QNLIMG-----LFLFLPVLLVRNSV-----LKGAIODR 419
DB 772 QCFTEFGAGELTKRLRYMFRSMLRQDVSWFDDPKNTGTALTRLANDAAQVGAIGSR 831
QY 420 V-----GLXQFGATPYTGLNANVNLFPVL-----RAVSDESQ 454
DB 832 LAVITOMIANNGTIIISFTIGMQLTILLAI--VPIIATAGVEMKMLSGQALKDKKL 889
QY 455 DGL-----YOKWQMLAYALHV-----LPESSVATMIF 482
DB 890 EGAGKRIATEAENRFTVSLIOEOKFEHMYAOSLOVRYRNSLRKAHIFGIFESTQAMMY 949
QY 483 SSV--CWTGLHBEVARFGFSALLAPHLIGEP--LTLVLLGIYONPNIVNSV----- 533
DB 950 FSYAGCF-----RFG--AYLVAHKLMSEFDVLLVFSAVVGAMAVGVSSRAP 995
QY 534 ---VALLSAGVLVSGFLNRIOEMPIPKIISYTFQKCYSELIVVNEFGALFTGSS 590
DB 996 DYAAKAKISAAHIH-----IIEKTPJ-----IDYSTBGLMPTLLEG--NVTGGEV 1039
QY 591 NVSVTTNPMCAFTOGI 606
DB 1040 VFNYPTRPDIPVLOGL 1055

```

RESULT 6
US-08-583-276-19

```

: Sequence 19, Application US/08583276
: Patent No. 5837536
: GENERAL INFORMATION:
: APPLICANT: McDonagh, Kevin T.
: APPLICANT: Niemiuis, Arthur
: APPLICANT: Tolstoshev, Paul
: TITLE OF INVENTION: IMPROVED EXPRESSION OF HUMAN
: TITLE OF INVENTION: MULTIDRUG RESISTANCE GENES AND IMPROVED
: TITLE OF INVENTION: SELECTION OF CELLS TRANSDUCED WITH SUCH GENES
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
: ADDRESSEE: Cecchi & Stewart
: STREET: 6 Becker Farm Road
: CITY: Roseland
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07068
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch diskette
: COMPUTER: IBM PS/2
: OPERATING SYSTEM: PC-DOS
: SOFTWARE: DM4.V2
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/583,276
: FILING DATE: 05-JAN-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/332,444
: FILING DATE: 31-OCT-1994
: APPLICATION NUMBER: 07/887,712
: FILING DATE: 22-MAY-1992
: INFORMATION FOR SEQ ID NO: 19:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1280 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE:
: DESCRIPTION: protein
: US-08-583-276-19

```

```

Query Match 7.2%; Score 240; DB 2; Length 1280;
Best Local Similarity 20.7%; Pred. No. 4e-16;
Matches 152; Conservative 108; Mismatches 230; Indels 246; Gaps 32;

```

```

QY 41 LHASYSVSHRYRPMWDTSCROMTROIKDVSILYVESGOIMCTIGSSGSKTLLDAMS 100
DB 397 VHFSPSRKEVK-----ILKGLNLKVGSGQVALVGNCGCKSTTVQLMQ 441
QY 101 GLRLRAGTFLGEVYVNGRALR--REQFQDCSYVLQSDTLSSLTVRETLHTYLLAIR 158
DB 442 -RL--YDPTGMSVSDODIRITINVRFLREITIGVSOEPVLFATTIAENIRY----- 490
QY 159 GNPSPQKVEAVAAE-----LSLHVADRLIGNSLGISTGERRRYSIAQLODP 211
DB 491 GREVNTDEIEKAVKEANADYFMKLPHKFTLVGENG--AOLSGGOKRIATIALVNP 549
QY 212 KVMLEDEPTGLDCTANOIYVLLVLELARNRRIYVLTIHOPRSELPOLFDTIALISFGE 271
DB 550 KILLDEATSLD--TESAVVQVALDKARKGRITIVIAH--RLSTVRNADVIAGFDGVI 606
QY 272 IFCTPRMELDFNDGCPCEHNSNPDFYMDLTSVDTOSKERIEIYS-----KRVQ 323
DB 607 VEKGNHDELM-----KEKGIYFKLVTMOTAGNEVELENADESKSEIDALE 652
QY 324 MIESAYKKSATCHTKLNIERMKHL-----KTLPMVPE-----KTRDSPG-- 363
DB 653 MSSNDSSSLRKSTRSRVSGSOADRKSLTKREALDESIPVPSFMRIMKLNLTMPYEV 712
QY 364 -----VESK--LGVLRL-----RVTRN-----LVNRKLAVITRL 391
DB 713 VGVECAIINGGLAPAFALIFSKIGVFTRIDDPETKRONSLFSLFLALGIIISFTFF 772

```

OY 392 QNLIMG-----LFLFFVLRNSV-----LKGAIODR 419
DB 773 QGTFGAGELTRKRLRYMFRSMLRODVSWFDDPKNTTGALTRLANDAAQVGAIGSR 832
OY 420 V-----GLYQVGAATPYTGMLNANVLPVL-----RAVSDQSO 454
DB 833 LAVTONTIANLGGITLISFYGMQLTLLAI--VPIIATAGVEMKMLSGOALKDKKL 890
OY 455 DGL-----YOKWOMLATALV-----LPSVATMIF 482
DB 891 EGAGKATEAIENFRVVSILTOBKFEHMTAOSLOVYRNSLRKAHIFGITFSFTQAMMY 950
OY 483 SSV--CYWTGLHPEVARFGFSALLAPHLIGEF--LTVLGLIYQNPINVSV-----533
DB 951 FSYAGCF-----RFG-----AYLVAKHLSFEDVLLVFSAYVFGAMAVQVSSFP 996
OY 534 ---VALLSTAGVLVSGFLNIOEMPPEKTIISFTFOKCYSEILVNEFYGLNFTCGSS 590
DB 997 DYAKAKISAAHIIM-----IIEKTP-----IDYSTEGIMPTLEG--NATFGEV 1040
OY 591 NNSVTTPMCAFTOGI 606
DB 1041 VFNYPTRPDIPVLOGL 1056

RESULT 7
5206352-4

Patent No. 5206352
Applicant: Ronlison, Igor B.; Pastan Ira H.; Gottesman,
Michael M.
TITLE OF INVENTION: COMPOSITIONS FOR CLONES CONTAINING DNA
SEQUENCES ASSOCIATED WITH MULTIDRUG RESISTANCE IN HUMAN CELLS
NUMBER OF SEQUENCES: 4
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/622,836
FILING DATE: 24-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 892,575
FILING DATE: 01-AUG-1986
APPLICATION NUMBER: 845,610
FILING DATE: 28-MAR-1986
SEQ ID NO: 4
LENGTH: 1280
5206352-4

Query Match 7.2%; Score 240; DB 6; Length 1280;
Best Local Similarity 20.7%; Pred. No. 4e-16;
Matches 152; Conservative 108; Mismatches 230; Indels 246; Gaps 32;

OY 41 LRAYSVSHRVPMMDITSCROQWTRQILKDVLSVSGQIMCITGSSGSKTTLIDAMS 100
DB 397 VHFSPSRKEV-----ILKGLNLAVOSQIVALVNSCGSKSTTVQLMQ 441
OY 101 GRLGAGTFLGEEVYNGRAL--REOFODCFSYVLOSPTLSSLVRETILHTALLAIR 158
DB 442 -RL--YDPLEGAVSDGODIRINVRFLREITIGVVSQEVLEFATIAENIR-----490
OY 159 GNPFSQKKEVNAE-----LSLSHADRLLINYSIGISGTERRRVSTIAOQLDOP 211
DB 491 GRENVTMDIEKAVEANAYDFIMKLPKFTDLVGERG--AQLSGGOKORIALARLVNRP 549
OY 212 KYMLFDEPTTGIDCETANOIVIVLVELARBNRIVLLTHOPSELFQFDKALISFBL 271
DB 550 KIILDEATSLD--TESAVVQVALDKARKGTTTIVAH--RLSTVRNADVIAGPDQVI 606
OY 272 IFCGTPAEMLDFNDGCGYCPREHNSPFEDYDLTSVDTSQEREIETS-----KRVQ 323
DB 607 VEKGNHDELM-----KEKGIYFKLVMTAGNVELENNADESKSIDALE 652
OY 324 MIESAYKSAICHKTIKRIEMKHL-----KTLPMVVP-----KTKDSPG--363
DB 653 MSSNDRSLLIRKSTRSVSGSOADRKLTSTKEALDESIPVSWFRIMKINLTWEMPYFV 712

OY 364 -----VESK-LGVLLR-----RVTRN-----LVNRKLAIVITRL 391
DB 713 VGVECAIINGLOPAFAIIFSKITIGVTRIDDPETKRQNSNLSFLFLALGIIISFITFEL 772
OY 392 QNLIMG-----LFLFFVLRNSV-----LKGAIODR 419
DB 773 QGTFGAGELTRKRLRYMFRSMLRODVSWFDDPKNTTGALTRLANDAAQVGAIGSR 832
OY 420 V-----GLYQVGAATPYTGMLNANVLPVL-----RAVSDQSO 454
DB 833 LAVTONTIANLGGITLISFYGMQLTLLAI--VPIIATAGVEMKMLSGOALKDKKL 890
OY 455 DGL-----YOKWOMLATALV-----LPSVATMIF 482
DB 891 EGAGKATEAIENFRVVSILTOBKFEHMTAOSLOVYRNSLRKAHIFGITFSFTQAMMY 950
OY 483 SSV--CYWTGLHPEVARFGFSALLAPHLIGEF--LTVLGLIYQNPINVSV-----533
DB 951 FSYAGCF-----RFG-----AYLVAKHLSFEDVLLVFSAYVFGAMAVQVSSFP 996
OY 534 ---VALLSTAGVLVSGFLNIOEMPPEKTIISFTFOKCYSEILVNEFYGLNFTCGSS 590
DB 997 DYAKAKISAAHIIM-----IIEKTP-----IDYSTEGIMPTLEG--NATFGEV 1040
OY 591 NNSVTTPMCAFTOGI 606
DB 1041 VFNYPTRPDIPVLOGL 1056

RESULT 8

US-08-665-259-25
Sequence 25, Application US/08665259
Patent No. 6028173
GENERAL INFORMATION:
APPLICANT: Landes, Gregory M.
APPLICANT: Burn, Timothy C.
APPLICANT: Connors, Timothy D.
APPLICANT: Dackowski, William R.
APPLICANT: Van Raay, Terence J.
APPLICANT: Klinger, Katherine W.
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: One Mountain Road
CITY: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665,259
FILING DATE: 17-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dugan, Deborah A.
REGISTRATION NUMBER: 37,315
REFERENCE/DOCKET NUMBER: IGS-9.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-3415
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1684 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

NAME: Plant, Thomas G.
REGISTRATION NUMBER: 35784
REFERENCE/DOCKET NUMBER: X9683
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
TELEFAX: 317-277-1917
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1307 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-395-246C-2

Query Match 6.9%; Score 229; DB 1; Length 1307;
Best Local Similarity 23.0%; Pred. No. 6,4e-15;
Matches 127; Conservative 110; Mismatches 188; Indels 128; Gaps 29;

QY 68 ILKDVSLVYESGQIMCIGSSGSKTLLDAMSGRLGR-AGTFLGEVY-----VNGRALRR 122
DB 407 VAHDLSCYIPACKTFAVGPSSGSKSTISLERFYDPVAGTMDGHDIGTLRLRLQ 466
QY 123 EGFQCFSVLSDTLSSLYRETLHTALAIRGNPGSFQ--KQYAV--MAE----- 174
DB 467 Q-----MSLVSEPRLEFAT-TIAENIRY-GIIGSFREKESYEIRKREVAARMANAHDF 519
QY 175 -LSLSHVADRLIGNSTIGSTGERRRVSIQAQLDPPKMLFDEPTGLDGMTANOIV 233
DB 520 IVALNGDYTNIESLSLQ--GOKORAIARAIIDPKILDLDETSALD--TKSEKLVQ 575
QY 234 LVELARRNRIVLTIHQPRSELFDLFDKIALISFGLFCSTPAEMLDFFNDGCPCE 293
DB 576 AALDKASKGRITIVIAH--RLSTIQKAVNIIVLANGQIEGPHHEIAD----- 622
QY 294 HSNPDFYMDLTVTOSKERELFETSKRVOMESAYKSAICHKTLKNIERKHLKTLPM 353
DB 623 ---RRGICDMVEAH-EIKKRISRSKRSQSOL-----LTNLSP-----KHNDM 661
QY 354 VPFKTKDSPG-----VF-----SKLGVLRRVYRNKLVAVITRLQNLINGFL 401
DB 662 TFFPFOKQYDQDESQIYSLSDADIGLHGEKRPVSRSLSLIMOPVEEAVSFMTL 721
QY 402 -----FVLNRSNVMLKGAIDRVAGLLYQFVGATPYTGMNANVLFVRLAV 448
DB 722 FKFLASFNRPEWPFLLGLCASILAGIOPSOAVLFAKAVSTLSPLPE---YKRLR-- 775
QY 449 SDQESDGLYKQWQMLVALHVLRFVSVAIMI-FSS--VCY-----WTGLHPEVARF 499
DB 776 ---HDANFCLMFIMIGIVSLVLYSVQGLTFAISSEKMYRRASQAFRVILHODISFF 830
QY 500 -----GYFSAALIA--PHLIGEFLLVILG--IVONPNIVNSV-----VALISI 539
DB 831 DQDENTGALATLTSAGTKELTG--ISSVTIGTILIVSVNLVASLGAVALVIGMKLALVCI 888
QY 540 AGV--LVGSGFLR 550
DB 889 SAVPALLMCGFVR 901

RESULT 14

US-09-134-001C-3641
Sequence 3641, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134, 001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064, 964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055, 779

PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3641
LENGTH: 273
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3641

Query Match 6.6%; Score 220.5; DB 4; Length 273;
Best Local Similarity 24.7%; Pred. No. 3.5e-15;
Matches 62; Conservative 52; Mismatches 82; Indels 55; Gaps 7;

QY 66 RQILNDVSLVYESGQIMCIGSSGSKTLLDAMSGRLGRAGTFLGEVYVNGRALRRQF 125
DB 31 KAVLEININKINQGEFLAIVGPGAGKSTLLKVI--LELPIQGEIIVDGPFPKGNKS 87
QY 126 QDCFSVILQ-----SDTLSSLYRETLHTALAIRGNPGSFQKQYAVM 172
DB 88 SLKISVSGKASAFNAGPPASVKEVLSGLTITKL-----FQFNKNDYQ-KYTVL 139
QY 173 AELSLSHVADRLIGNSTIGSTGERRRVSIQAQLDPPKMLFDEPTGLDGMTANOIV 232
DB 140 ERLNISHLI-----NKNIAELISGGQOQRYLIARALISPSVVLDEPTNGIDAKHVSFY 194
QY 223 VLLVELARRNRIVLTIH-----QRSELFDLFDKIALISFGL 271
DB 195 DTLDKLKKQGITIILVTIDIGVADTAEVACLNRHLPHGSTEAFKSLIDEVEISK----- 250
QY 272 IFGCTPAEMLD 282
DB 251 -IYGHPIQFVD 260

RESULT 15

US-08-665-259-27
Sequence 27, Application US/08665259
Patent No. 6028173
GENERAL INFORMATION:
APPLICANT: Landes, Gregory M.
APPLICANT: Burn, Timothy C.
APPLICANT: Connors, Timothy D.
APPLICANT: Dackowski, William R.
APPLICANT: Van Raay, Terence J.
APPLICANT: Klinger, Katherine W.
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESS: GENZYME CORPORATION
STREET: One Mountain Road
CITY: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665, 259
FILING DATE: 17-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dugan, Deborah A.
REGISTRATION NUMBER: 37,315
REFERENCE/DOCKET NUMBER: 1G5-9.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 1457 amino acids

Mon Jul 28 09:43:29 2003

us-09-989-981a-6.rai

```

; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-665-259-27

```

Query Match	6.6%;	Score 220;	DB 3;	Length 1457;
Best Local Similarity	32.1%;	Pred NO	7.3e-14.	

best local similarity 32.18; Pred. No. 7:3e-14;
Matches .69; Conservative 37; Mismatches 91

Matches 69; Conservative 37; Mismatches 91; Indels 18; Gaps 8;

OY 79 GCMCIISSSGSKTTLIDAMSCRLGRACFTFLGEVYVNGRALRRE--QFDCFSYLQSD 136
| : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : :
Db 1104 GEFGGLGVNAGACTSTFKMLTDDESTG--GRAEVNHSLVDLLDYOOSSLCCPDP 1160

Db 1104 GECFGLGVNGAGKTSYFKMLTDESTWG--GEAFVNGHSYLKDLLQVQSSLGCCPQFD 1160

137 TLSSLTVRETLH-YTALLAIRGNPGSPQOK-VEAVMAELSLSHVADRLIGNYSLGIGS 194

Db 1161 VPVDELTAREHLDYTRLRCL----PMKDEQNVKMALEKLELTTRYADKPAGTY-----S 1211

Oy 195 TGERRRVSIAQQLDDPKVMEDEPTTGLDGMTANQIVLVVELARNRIVLTIHQPRS 254
|:::| |:::|

D_b 1212 GGNKRKLTALALIGYPAFIFLDEPTTGNDPKARRELMNLIDLITGRSVLTSHS-ME 1270

QY 255 ELFQLEFDKIAILSFGELIECGTPAEMLDPEFNDGCGY 289

Db 1271 ECEALCTRLAIVNGRLHCLGSIQHLKNRFGD-GY 1304

Search completed: June 11, 2003, 09:14:17
job time : 22 secs

Page 9

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 09:09:57 ; Search time 22 Seconds

(without alignments)
2844.705 Million cell updates/sec

Title: US-09-989-981a-6

Perfect score: 3326 1 MGDLSLTPGSGMGLOVNRG.....PALVILGIYFKIRDLISR 651

Sequence: BLOSUM62

Scoring table: Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 10%
Listing first 45 summaries

Database :

1: PIR73:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	649.5	19.5	725	2	ABC transporter-11
2	627	18.9	1294	2	ATP-dependent perme
3	621	18.7	1049	1	ABC transporter-11
4	614	18.5	720	2	probable ABC trans
5	610.5	18.4	725	2	ABC transporter-11
6	608	18.3	708	2	ABC transporter-11
7	602.5	18.1	687	1	white protein - fr
8	597	17.9	687	1	protein F12M6.17
9	595.5	17.9	658	2	hypothetical prote
10	591.5	17.8	740	1	probable ATP-bind
11	590.5	17.8	638	2	white homolog - hu
12	589.5	17.7	646	2	probable ABC trans
13	584	17.6	755	2	probable ABC trans
14	583.5	17.5	687	2	hypothetical prote
15	578.5	17.4	646	2	ATP binding cass
16	575.5	17.3	649	2	probable ABC trans
17	572	17.2	739	2	ATP transporter-11
18	571.5	17.2	609	2	probable ABC trans
19	564	17.0	662	2	ABC transporter-11
20	554	16.7	559	2	ABC transporter-11
21	552.5	16.6	635	2	protein COSD10.3
22	545.5	16.4	608	2	hypothetical prote
23	540	16.2	678	2	hypothetical prote
24	539	16.2	577	2	hypothetical prote
25	529.5	15.9	633	2	hypothetical prote
26	527.5	15.9	659	2	hypothetical prote
27	517.5	15.6	610	2	hypothetical prote
28	517.5	15.6	639	2	protein C10C6.5
29	514.5	15.5	705	2	probable ABC trans

30	504	15.2	695	2	T21109	hypothetical prote
31	497	14.9	547	2	T31543	hypothetical prote
32	488	14.7	737	2	T46101	ABC transporter-11
33	464	14.0	1450	2	A84780	probable ABC trans
34	456.5	13.7	1426	2	T30567	ATP-binding cass
35	454.5	13.7	675	1	FYFPR	ATP-binding cass
36	435	13.1	668	2	S55023	ATP-binding cass
37	429	12.9	1469	2	H96622	ATP-binding cass
38	425.5	12.8	1443	2	T02491	ATP-binding cass
39	425	12.8	1451	2	B86286	ATP-binding cass
40	421	12.7	1450	2	T45888	ATP-binding cass
41	420	12.6	1435	2	D96693	ATP-binding cass
42	418.5	12.6	1420	2	T02644	ATP-binding cass
43	417.5	12.6	1619	2	T30541	ATP-binding cass
44	413	12.4	1413	2	G84790	ATP-binding cass
45	410	12.3	1177	1	D71416	ATP-binding cass

ALIGNMENTS

RESULT 1		ABC transporter-like protein - Arabidopsis thaliana	
T47652		N:Alternate names: protein T26112.10	
C:Species: Arabidopsis thaliana (mouse-ear cress)		C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000	
C:Accession: T47652		R:Montfort, A.; Casacuberta, E.; Puidomenech, P.; Meyers, H.W.; Lemcke, K.; Mayer, submitted to the Protein Sequence Database, February 2000	
A:Reference number: Z24471		A:Accession: T47652	
A:Status: preliminary		A:Molecule type: DNA	
A:Residues: 1-725 <MON>		A:Experimental source: EMBL:AL132954	
A:Genetics:		A:Map position: 3	
A:Note: T26112.10		C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1;	
Query Match		Best Local Similarity 19.5%; Score 649.5; DB 2; Length 725;	
Matches 182; Conservative 124; Mismatches 246; Indels 67; Gaps 15;			
QY	33	PEPHSIGILHASYSHRYRPMWMDITSCROQTROIADVSLYVESGOIMCLGSSGSK 92	
DB	68	PVPYVILNFRNLDYDTLRR-----RFSRQNGVKTLDDVSGEASDGLAVLAGSGACK 123	
QY	93	TTLDAMSGRLGACFTLGEVYVNG-RALRREPODCFSYVQSPTLSLTVRETIHYT 151	
DB	124	STLIDALAGRAVE-GSLRGSVTLNKEKVLQSLKLVISAYVQDDLPMLVKEETMFA 182	
QY	152	ALLAIRRG-NRGSFQKYEAYVAELSLSHVADRLIGNSLGISIGERRRVSIAAQLD 210	
DB	183	SEFRPLRSLSKKKEVEREALDQLGRNAATVGDGEGKRGSGERRRVSIGIDIHD 242	
QY	211	PKVMEDEPTTGIDCMATNOIVLVLVELARRNRIYVLTIHOPRSLEFQFKIALISPE 270	
DB	243	PVLEFLDEPTSGLDSTNAFMVYQVLAQSGSIVTMSIHOSARIVELDLILISRGK 302	
QY	271	LIFCGTPAEMLDFFNDGCPCEPHSNPDEFYMDLTVTQSKREIETS----- 319	
DB	303	SVFNQSPASLPQGFDFSRPIPEKENISEFALDY-----RELQSGNEKTLVDPN 354	
QY	320	-----KRYQIESAVK-----KSAICHTLT-KNIERNKHLKTLPMVPFKTRD 360	
DB	355	EKMDOONKISLQSAVQTKLDDQRLSLSLKEAIVNASVSGKLVSGSSRSNPTSMETVSSA 414	
QY	361	SPGVESKGLVLRVTRNLVKNKLAVITRLLQNLGFLFFVLRVSNVUKGAIQDRV 420	
DB	415	NPSLFEPR-ILAKRIKMKWITRPELAVGTRIAVAVTGC-LTAVYWKLDHPTPRGA-QERTL 471	

QY 421 GLTYQVQATPTTGMLNANVLPVLRAVSDQDGLYOKMOMLAVLHVPSSVATM 480
 Db 472 -TLFAVYPMFEYCCIDNVPVFIQERYIFLRETHNARSSSVYSLVSLPOLAPSL 530
 QY 481 IFSSVCYTGIGLHPEVARGYSALLPHLIGELTVLIGYONPNIVS-VYALLSI 539
 Db 531 VFSATFTVTGLSGGSEGFVYCLITIASFWSGSSVTFISGV--PRIMICMYSTYL 588
 QY 540 AGVLVSGFLNIOEMPIPKIISYTFQKYCELIIVNVEFYGLNFTGSSNSVATPM 599
 Db 589 AYCLLSGFYVNRDIRPPTWTFHISLTKIPYEVANLNEF-----DDPS 633
 QY 600 CAFTOGIOFIETKCPGATS 618
 Db 634 RCFVGVQVDFDSTLLGVS 652

RESULT 2

Probable membrane protein YOL075c - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein O1125; hypothetical protein O1130; hypothetical
 C:Species: Saccharomyces cerevisiae
 C:Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 19-Apr-2002
 R:Alexandrak, D.; Katsoulou, C.; Tzermia, M.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S66756
 A:Accession: S77690
 A:Molecule type: DNA
 A:Residues: 1-1294 <ALE>
 A:Cross-references: EMBL:Z74816; MIPS:YOL075c
 A:Note: this is a revision to the sequence from reference S66756
 A:Accession: S66767
 A:Molecule type: DNA
 A:Residues: 1-179 'TTRTGVLYVKRED' <ALM>
 A:Cross-references: EMBL:Z74816
 A:Experimental source: strain S288C
 A:Note: this sequence has been revised in reference S77690
 A:Accession: S66768
 A:Molecule type: DNA
 A:Residues: 200-1294 <ALF>
 A:Cross-references: EMBL:Z74817
 A:Experimental source: strain S288C
 A:Note: this sequence has been revised in reference S77690
 C:Genetics:
 A:Cross-references: SGD:S0005435
 A:Map position: 15L
 A:Note: YOL075c
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
 C:Keywords: ATP; nucleotide binding; P-loop; transmembrane protein
 F:45-263/Domain: ATP-binding cassette homology <ABC1>
 F:62-69/Region: nucleotide-binding motif A (P-loop)
 F:316-352/Domain: transmembrane #status predicted <TM1>
 F:469-485/Domain: transmembrane #status predicted <TM2>
 F:496-512/Domain: transmembrane #status predicted <TM3>
 F:606-622/Domain: transmembrane #status predicted <TM4>
 F:710-916/Domain: ATP-binding cassette homology <ABC2>
 F:727-734/Region: nucleotide-binding motif A (P-loop)
 F:1042-1058/Domain: transmembrane #status predicted <TM5>
 F:1125-1141/Domain: transmembrane #status predicted <TM6>
 F:1177-1193/Domain: transmembrane #status predicted <TM7>
 F:1269-1285/Domain: transmembrane #status predicted <TM8>

Query Match 18.98; Score 627; DB 2; Length 1294;
 Best Local Similarity 31.7%; Pred. NO. 1,1e-38;
 Matches 181; Conservative 106; Mismatches 228; Indels 56; Gaps 19;

QY 65 TROLKDVSLYVSGOIMCTIGSSGKTLIDAMGRGRATFL-----GEVYVNGRA 119
 Db 706 TKELIOSVNAIFKPGMINAIMGPGSGKSSLLNLISGR-L-KSSVFAKFDPTSGSIMPNID 764

QY 120 LRREFQDCFSYVLQSDT-LISLSLRETHYTLALTRGNGPSFOKVEANMAELIS 178
 Db 765 VSELMFKVNCYSVSDDDHLAALTAKTLLYAAALRHHLTEERMERTNLRSGLK 824
 QY 179 HVADRLIGNSLGISTGERRRVSIAQLQDPYMLDEFPTGLDCMTAQIVLVEL 238
 Db 825 HCENNIIGNNEVKGISGGERKRRVMGVQLNDPILLDELFTSGLDSTSTLIELEKL 884
 QY 239 AR-RNRIVVLTIRHPRSLFOLFKAIALS-FGLIFCGTPAEMLDFFNDGCPPEHSN 296
 Db 885 CREGKTLITIRHPRSELFRPGVNLILASGRTAFNGSPDEMAVTEELGYNCPSFTN 944
 QY 297 PFDFYMDLTSVDIOSKEREIETSKRVOMIESAYKSAICHTIKMEKHLKTPMPFP 356
 Db 945 VADFFDLISVNTONEQNEISSRAVEKILISAMK-----AMDN-ESLSPTPISEK 994
 QY 357 KTKDSPGVSKLGVLLRRETNLY-----RNKLAVTTLNLNIGLEFL 401
 Db 995 QOYQSEPFTEYSEVVR-PANVLAVLVNVRQFTTTRRSPDSLMAIADIPGLGVFA 1053
 QY 402 FFVLRVSNVILKGAIDQRVGLLYQFVGTPTGMLNANVLPVLRAVSDQDGLYQKW 461
 Db 1054 LFFAPVKNHTY--SISNRLGIAQEST-ALYFVGMGLNACYPTEEDYEEYNVYGIA 1110
 QY 462 QMLATYALHVPFVSVAFTMIFSSVCYWTGLGHPVARGYSALLAPHLI---GEFLTL 518
 Db 1111 PFLAVMTLELPISALASVLAFTVLAAGL-PRTA--GNFPAVYCSFIYTCGERLGI 1167
 QY 519 VLIGYONPN-IVNSVALLSIAGVLVSGFLRNIOEMPIPKIISYTFQKYCELIIV 577
 Db 1168 MTNFFRPRPGVAVNCISIIISIGOMGSLMST---GMSRYLKGFNTLNPVGYTSMIIN 1223
 QY 578 NEFTG-LNFTC--GSSNVSVTTPMCAFTOG 605
 Db 1224 FAFPGNLKICEDCGKNSDGT---CEFANG 1250

RESULT 3

ATP-dependent permease ADP1 precursor - yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein YCR011c; protein YCR105
 C:Species: Saccharomyces cerevisiae
 C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
 R:Goffeau, A.; Purnelle, B.; Skala, J.
 submitted to the Protein Sequence Database, March 1992
 A:Reference number: S19420
 A:Accession: S19421
 A:Molecule type: DNA
 A:Residues: 1-1049 <GOF>
 A:Cross-references: EMBL:X59720; NID:g1907116; PIDN:CAA42328.1; PID:g1907154; GSPDB:
 R:Purnelle, B.; Skala, J.; Goffeau, A.
 Yeast 7, 867-872, 1991
 A:Title: The product of the YCR105 gene located on the chromosome III from Saccharom
 A:Reference number: S40914; MUID:92160395; PMID:1789009
 A:Accession: S40914
 A>Status: not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-1049 <PUR>
 R:Skala, J.; Purnelle, B.; Goffeau, A.
 Yeast 8, 409-417, 1992
 A:Title: The complete sequence of a 10.8 kb segment distal of SUR2 on the right arm
 K genes.
 A:Reference number: S25353; MUID:92327849; PMID:1626432
 A:Contents: annotation
 A:Genetics: annotation
 A:Gene: SGD:ADP1; MIPS:YCR011c
 A:Cross-references: SGD:S0000604; MIPS:YCR011c
 A:Map position: 3R
 C:Superfamily: ATP-dependent permease ADP1; ATP-binding cassette homology
 C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop; transmembrane protein
 F:1-25/Domain: signal sequence #status predicted <SIG>

guss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: C84423

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-725 <STO>

A:Cross-references: GB:AE002093; MID:94262239; PIDN:AD14532.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g01320

A:Map position: 2

Query Match

Best Local Similarity 29.7%; Pred. No. 8.6e-38;

Matches 166; Conservative 111; Mismatches 244; Indels 37; Gaps 11;

```

OY 51 VRP-----WMDITSC-----ROMTQIILKDVSLVYESQIMCIGSSSGKTTLDAMSGR 102
Db 65 IRPPTIMRNITQSLSDKSSKSVFLLKNVSGEAKPRLLAIMGPSGGKTTLLNLVAGQ 124
OY 103 LGRAGT--FLGEVYVNGRATLRBOFCFVSVLQSDTLSSLTIVRETLHTALLAIR-G 159
Db 125 LSLSPRLHLSGLEVNCKRPSSSKAYK--LAFVROEDLFFSOLTVRETLSPAELQLEIS 182
OY 160 NPGSFQKVEAVNAELSLSHVADRLIGNYSIGISTGERRVSIATLQDPPKMLPDEP 219
Db 183 SAEERDEYVNNLLKFLGLVSCADSCVGDAAVIRGISGCKRRLSLACGLINSPVIFRDEP 242
OY 220 TTGLDCTANOIYVLLVELARNRIVVLTTHORSELPFLPKATLISFEBELFCGTPA- 278
Db 243 TTGLDAFOAKKVMETLOKLAODGHVYICSIHOPGSAVYAKFDDIVLLTEGTLVYAG-PAG 301
OY 279 -EMLDFFNDGCPCEPHSNPFDFYMDLTVDTQSKEREIEFSKRVOMIESAY--KKAIC 335
Db 302 KEPLTYFGNGFLCPENVAEFLADLIDSYSSSEVYSQKRAHALVDAFSORSSVL 361
OY 336 HKTLKNIERKHLKTLPMVPFKTKDSPGVESKLGVLRLV-----TRNLVANKLAVI 387
Db 362 YATPLSMKEETKNGMPRRKAIYERTDGMWRQFFLLKRAMQASRDGPINKYARMSVA 421
OY 388 TRLLQNLINGLFLFPLRVRNVLGAIDRRGLLYQFGAIPYTGMLNANVLPVLR 447
Db 422 SA-----VIFSGVWRMKSQTSIDRRMGLLOVAALNTMAALITTCVGFPERA 471
OY 448 VSDQSDGLYQKQOMALATLHLPLFSVATMIFSSCYWTGLGHEPARFGVFSALL 507
Db 472 IYDRERSKGSYSGPYLLSTKTIAEIPGAFLPMFGAVLIPMARLNTLSRFKFGCIYV 531
OY 508 APHLIGEPFLTLVLGIYQNDNIYNSVALLISAGVLVSGFLRNIOEMPPIFKIISYTFP 567
Db 532 VESFAASAMGLTFGAVVPSTEAMAVGPSIMTVFIYFG--GYVNADWTPIIFMIPRASL 590
OY 568 OKYCEELIVVNEFYGNF 585
Db 591 IRMAFGGLCINERFSGLKF 608

```

RESULT 6

T47650

ABC transporter-like protein - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000

C:Accession: T47650

R:Newes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.

submitted to the Protein Sequence Database, April 2000

A:Reference number: Z24470

A:Accession: T47650

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-708 <MEM>

A:Cross-references: EMBL:AL132970

A:Experimental source: cultivar Columbia; BAC clone T15C9

C:Genetics:

A:Map position: 3

A:Note: T15C9.110

C:Superfamily: *Arabidopsis thaliana* probable ATP-binding cassette protein F12U6.1; A

Query Match

Best Local Similarity 27.5%; Pred. No. 1.3e-37;

Matches 171; Conservative 136; Mismatches 256; Indels 58; Gaps 15;

```

OY 18 NRGSSLEGAPA--TAPEHSLGILHASYSVSHRVPMMDITSCROMTROIILKDVSLY 75
Db 41 NAPTQHLIDLAPAEATRSVPFLSFNNLSTNNVLR--R--EDFSRRKATASVKTLLDITGE 98
OY 76 VESQIMCIGSSSGKTTLDAMSGRLGAGFTFLGYYNG--RALRREPQDCFSYVLQ 134
Db 99 ARDEILLAVLGSSGAGKSTLIDLAGRYAE--DSLKGTIVTNGEKVLSRLKTVASAYVQ 157
OY 135 SDTLSSLTIVRETLHTYALLAIRGNNGSFO--KKEVAVNAELSHVADRLIGNYSIGI 193
Db 158 DDLFLPMLTYKETIMFASERLPRSLPKSKMERVETLIDQLGRNADTVIGDEGRGV 217
OY 194 STGERRRVSIATLQDPPKMLPDEPTGLDCTANOIYVLLVELARNRIVVLTTHQPR 253
Db 218 SGERRRVSIATLQDPPKMLPDEPTGLDCTANOIYVLLVELARNRIVVLTTHQPR 277
OY 254 SELFQDPKATLISFGLICGTPAEMLDFNDGCPCEPHSNPFDFYMDL-----TSV 307
Db 278 ARIIGLDRLLILSHGKSVNGSPVSLPSEFSGRPIPERKENTFERALDVIRESGSE 337
OY 308 DT-----OSKREIETSKRVOMIE-----SAVKAICHKTLKNIERKHLKTL 352
Db 338 CTNDLVEFNKMOONQARATTSRVSILKALIASVSKGLVSGSSAANISMETVSSYA 397
OY 353 MVEPKTKDSPGVESKLGVLRLVRLVNRKLAIVTRLLQNLINGLFLFPLRVSNVL 412
Db 398 NRP-----LAEPFLAKRYIKMIRPELIGRIGVAVMTGLLAVYRL--DNTF 447
OY 413 KGAIQDRVGLLYQFGAIPYTGMLNANVLPVLRANVSDQSDGLYQKQOMALATLHL 472
Db 448 RGA-QERMG--FFAFGSTMTEYCCADNIPVIOERYITFLRBTINAVRTSSYSVALVSL 505
OY 473 PFSVATMIFSSCYWTGLGHEPARFGVFSALLPHLIGEPFLTLVLGIYQNDNIYNS 532
Db 506 POLLALSIANAATFTVGLSGLESEFFYCLIIYAIFMSGSSIVFISGL--PVMMS 563
OY 533 -VALLISAGVLVSGFLRNIOEMPPIFKIISYTFQKCYSELIVVNEFYGNLFTGSSN 591
Db 564 YMTIAYLISYCLLGEFYINRDIRIPYMIWEHYISLKYPEAVLINF-----612
OY 592 VSYTNPMAFQIOGIEKT 612
Db 613 ----DPSRCFVKGVQVFDGT 629

```

RESULT 7

FYFPM

white protein - fruit fly (*Drosophila melanogaster*)

C:Species: *Drosophila melanogaster*

C:Date: 31-Dec-1990 #sequence_revision 17-Feb-1995 #text_change 19-Jan-2001

C:Accession: S08635; S07263; S10240

R:Pepling, M.; Mount, S.M.

Nucleic Acids Res. 18, 1633, 1990

A:Title: Sequence of a cDNA from the *Drosophila melanogaster* white gene.

A:Reference number: S08635; MUID:90221897; PMID:2109311

A:Accession: S08635

A:Molecule type: mRNA

A:Residues: 1-687 <DEP>

A:Cross-references: EMBL:X51749; MID:g8825; PIDN:CAA36038.1; PID:g8826

R:O'Hare, K.; Murphy, C.; Lewis, R.; Rudin, G.M.

J. Mol. Biol. 180, 437-455, 1984

A:Title: DNA sequence of the white locus of *Drosophila melanogaster*.

A:Reference number: S07263; MUID:85134865; PMID:6084717

Db 557 SORNS-NIGIMASPIVGVRLGPFILMYR 564

RESULT 9

T31958

hypothetical protein F02E11.1 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #extl_change 31-Jan-2000

C:Accession: T31958

R:Ravello, A.; Scheet, P.

submitted to the EMBL Data Library, July 1997

A:Description: The sequence of *C. elegans* cosmid F02E11.

A:Reference number: 221104

A:Accession: T31958

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-658 <FAV>

A:Cross-references: EMBL:AF016661; PIDN:AB66050.1; GSPDB:GN00020; CESP:F02E11.1

A:Experimental source: strain Bristol NZ; clone F02E11

C:Genetics:

A:Gene: CESP:F02E11.1

A:Map position: 2

A:Introns: 115/3; 158/3; 214/3; 330/3; 368/2; 448/3; 525/1

C:Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match

Best Local Similarity 17.9%; Score 595.5; DB 2; Length 658;

Matches 165; Conservative 121; Mismatches 255; Indels 67; Gaps 11;

72 VSLVESGOIMCITLSSGSGKTTLLDAMSGRLGAGTFLGYVNGRALRPOFDCESY 131

79 VSGVAPGEVLAIMSGSGKTTLLDAMSGRLGAGTFLGYVNGRALRPOFDCESY 138

132 VQSDPTLLSLVRETLLTALAIRGNPSGFOKVEVMAELSLSHADLLG-NSL 190

139 VQVVDLFCGTILVRLQTLTAMRKKNATVQOKMEVEVNLKDMNLTDQNTLISPMN 198

191 GISTGERRRVSIAQLLQDPKMLDEPTGLDCTANOIVLVLELARRRIVLTH 250

199 KGISIGEKRLAFACEILDPKILFCDEPTSLDPMASEVVRALLDLANKKTIIVLH 258

251 OPRSLFOLFQKIALISFGEILFCGTPAEMDLFNDCC--YCPESHNPFDYMDLTVD 308

259 QPSSTVFRRFHVCMATGKTVYHGAVDRLCPFPDLGSPDFVPSYPADPVMSEIST- 317

309 TOSKREIETSKRVOMISAVKSAICHKTLKNIEMHKLTPVWPPTKDSPEVSKL 368

318 --SPETBOEDVTRIRIYLHEIYONSDIGTOMLK-----KTRTAVDGEGY 359

369 G-----VLLRVRTNLVNRKLAVITRLLQNLIMGLFLFVLRYRS 409

360 GDEDDGESRYNSTFGTOFEILKRSKLTFRDPLLRPAQLLATIIVGVNRYE- 418

410 NTLKG-AIODRGLLXQVGGATPYTGMLNANVLFVYLAVSDOESODGLYOKOMMLAVA 468

419 --LKPQTIONEGVYVNCARDMTFLFPSSVAVITSELVPLREKHSIYSEAVEFLAKS 476

469 LALVFSVVAATMIFSSVCWTLGLHPVARFGYFSAALLAPHLJEFILVILGIVQPN 528

477 LAELPQYTLIPRIYGTIYMAAGLVASTSLVLFVFCITTLTAVASIAIYGACIFGEG 536

529 IYNSVALLSTAGVLVSGFLANIQEMPIPKIISYTFQYKXCEILVYNEFYGLNFTCG 588

537 LVTFTPMF-VLPMLVFGGFYVANSIPYVYQVSVFVSMFKHGFALANOMKEIDKSG 595

589 SSNSVVTNPMCAFTQGTQTEKCP-----GATSRFTMNFILYSLFALVYL 637

596 CDLI---NPLNATTGT-----CPASDGGIILTRRIDPPLVYANVILTFMSEFVYRII 645

638 GIVVEKIR 645

646 GIVVALKIR 653

RESULT 10

T02567

probable ATP-binding cassette protein T16B24.1 - *Arabidopsis thaliana*

M:Alternate names: protein F12L6.1

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #extl_change 02-Mar-2001

C:Accession: T02567; T00545; C84816

R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.

submitted to the EMBL Data Library, August 1998

A:Description: *Arabidopsis thaliana* chromosome II BAC T16B24 genomic sequence.

A:Reference number: T14679

A:Accession: T02567

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-740 <KROD>

A:Cross-references: EMBL:AC004697; NID:g3402671; PIDN:AAC28975.1; PID:g3402672

A:Experimental source: cultivar Columbia

R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.

submitted to the EMBL Data Library, July 1998

A:Description: *Arabidopsis thaliana* chromosome II BAC T12L6 genomic sequence.

A:Reference number: T14168

A:Accession: T00545

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-362 <KROD>

A:Cross-references: EMBL:AC004218; NID:g3355463; PIDN:AAC27826.1; PID:g3355464

A:Experimental source: cultivar Columbia

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Taylor

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A:Reference number: A04420; MUID:20083487; PMID:10617197

A:Accession: C84816

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-740 <STO>

A:Cross-references: GB:AE002093; NID:g3402672; PIDN:AAC28975.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g39350; T16B24.1; F12L6.1

A:Map position: 2

C:Superfamily: *Arabidopsis thaliana* probable ATP-binding cassette protein F12L6.1; A

F:110-310/Domain: ATP-binding cassette homology <ABC>

Query Match 17.8%; Score 591.5; DB 1; Length 740;

Best Local Similarity 27.5%; Pred. No. 2.4e-36;

Matches 191; Conservative 123; Mismatches 267; Indels 113; Gaps 19;

35 PHSGLIMASYSVSRVPMWD-----ITSCROQTRQLKDVSLVESGQ 80

64 PVLSEFDMULTYVNS--VRPKLDFRNLFPRRRTDEPEIAOTARPKTKTLNINISGSTRDGE 121

81 IMCILGSSGSKTTLTLDAMSGRLGAGTFLGYVNGRALRPOFDCESYVLSDTLLS 140

122 IMAVLGAGSGSKTLLDILANRIAK-GSLKGTIVKNGETLDSRLMKVISAIVMODLLFP 180

141 SLTVRETLHYTALLAIRGNPSGFOK-FVEAVMELSLSHVADRILGNSLGIETGERR 199

181 MLYEETLMFAEFPLPSLPRKSKKILVQALIDGLIRNAKTIIGEGHRSIGSGRR 240

200 RVSTAAQLOLPKWLDEPTGLDCTANOIVLVLELARRRIVLTIHQPSSELPOL 259

241 RVISGIDILHDPILFLDEPTSGDSTSAFVYKVLIAOSGSIVIMSHQPSHRVYGL 300

260 FDKIALISFGEILFCGTPAEMDLFNDCCYCPESHNPFDYMDLTVDQSKERE----- 315

301 LDRILFLSRGHTVYSGSASLPRFTTEGSGIPENENTETRALDLI-----RELGSAG 354

316 -----IETSKRVQITE-----SAVYKSAICHKTLKNIEMK----- 346

355 GTRGLIERNKMKQKQSNROPPLTPSSPYPNLTLEATAAISRSCKLVSGGGSVAHG 414

QY 347 ----HLKTLPMVPEKTKDSQVFSKGLVLRRTVNLVKNKLAIVTRLLQNLMLGELLF 402
DB 415 GATTTTTLTAVAFANP-----MMIEIKLSKRSMLNSRQPELFCIRIASVITG-FILA 469
QY 403 FVLRRSNVLRKCAIDRRGLLQYFGCAPPYGMUNAVNLFVPLRAVSQDESDDGLQKQK 462
DB 470 TVFRLDNPSPK-VQERLG-FEAFAMSTMETTCACALPEVLOERYIFMHEFYANVARRSS 527
QY 463 MMLAVALHVLPSVAVATMIFSSVCYWTGLHPEVARFGESALLAPHLIGELFLVILG 522
DB 528 YLASHAIVSFPSELIFLSAFAATTYTAVAGLOGGLGLFLYCLIIILASWSSGFVTFISG 587
QY 523 IVQNNIV---NSVALLSIAGVLVSGFELRNIOEMPPIPKIISYTFQKCYCEILVYNE 579
DB 588 VV--PSVMLGYTIVAIL-AVFLFSGFEFINRNRIPDYIMFHMVMSLVKRYEAVLQNE 643
QY 580 FYGLN--FTCG-----SSNVSTTNMCAFTQG----- 605
DB 644 FSDATKCFVRGVQIFDNTPLSELPEVMKLLGLTVSKSLGVYISSTTCLTGLSDILROOG 703
QY 606 -IOPLEKTCGATSRFTNPLFLYSFIPALVILG 638
DB 704 VVOLSKMNCFLITVAFGFFFLIFYE---TLLIG 734

RESULT 11

G02068
white homolog - human
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 02-Feb-2001
C:Accession: G02068
R:Crump, J.M.; Tiller, G.; Fletcher, J.A.; Lux, M.; Raab, E.; Goldenson, D.; Archiniegas, submitted to the EMBL Data Library, August 1995
A:Reference number: H00769
A:Accession: G02068
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-638 <CRO>
A:Cross-references: EMBL:U34919; NID:g1314276; PIDN:AAC51098.1; PID:g1314277
C:Genetics:
A:Gene: white
C:Superfamily: fruit fly white protein; ATP-binding cassette homolog
C:Keywords: ATP; nucleotide binding; P-loop
F:61-253/Domain: ATP-binding cassette homology <ABC>
F:78-85/Region: nucleotide-binding motif A (P-loop)

Query Match 17.8%; Score 590.5; DB 2; Length 638;
Best Local Similarity 26.3%; Pred. No. 2.3e-36;
Matches 164; Conservative 142; Mismatches 26; Indels 51; Gaps 14;

QY 44 SYSVSHRRPMMWDTSCROOQWROIILKDVSLYVESGOIMCIGSSSGKTTLLDAMGRL 103
DB 43 SYSVPE--GRRW-----RKGKYLTLGISEKENSSELVAMIGSGAGKSTLMLNLGY- 94
QY 104 GRAFTLGEYVYNGRALRRQFQDCFSYVLOSDTLLSLVRETLHTTALLAIRGNPGS 163
DB 95 -RETMGMGAVLNLPRDLRCFRKVSXYIMODMLLPHLVYQEMAMSAHLKLOEKDEGR 153
QY 164 FQKVEVMAELSLSHVADRILGNYSLGISTGERRRVSIAQLOLDPKVMLEDEPTTGL 223
DB 154 -REVMKELTALGLSCA-----NTRTGSLSGGRKRLAIALLEVNRPVMEFDEPISGL 207
QY 224 DCAHTANIYVLVELARNRIVITIHQPSSELPQLDKAIALISFGLICGTPAEMIDF 283
DB 208 DSAACFOVSLMGLAGGSRISICTTHQPSAKLEFLDOLYVLSQGCYVRGKVCNLPY 267
QY 284 FNDGVPCEPSPHPDFYMDLTSYDQSKREIETSKRYOMISAVYKSA-----I 334
DB 268 LRDGLNCPYTHNPADPVMEVASGEYDQSKRLVRAVRBGMCSDBKRDGLGDAEVNPL 327
QY 335 CHHTLKIEEMKHLKTLPMVPEKTKDSQV-----FSKGLVLRRTVNLVKNK 384
DB 335 CHHTLKIEEMKHLKTLPMVPEKTKDSQV-----FSKGLVLRRTVNLVKNK 384

DB 338 WHPSEEVAKQTRKLGK-----RKDSSMECHSFSASCLTQPCILFRKFLISMDSV 381
QY 385 AVITRLQNLMLGELLFEEVLRVRSNVLKGAIQDRVGLLYQVCAPTPTGMNAVNLFPV 444
DB 382 LTHLSTHIGLIGLILGLYLGIGNEARK--VLSNSGLFFSMELPMAALPVLTPPL 439
QY 445 IRAVSDOSODLYOKOMOMLALVAVLHVPESVAVATMIFSSVCYWTGLHPEVARFGES 504
DB 440 EMGVFLRHLNWSLAKYVLAKTADVPFOIMEFVACSTYVMTTSSQPSDAVAVLFP 499
QY 505 ALLAPHLIGELFLVILGIVQNPVINSVALLSIAGVLVSGFELRNIOEMPPIPKIISY 564
DB 500 LGTWSLVAQSLGL-LIGAASLQVAFVGVTAIPVLLPSGFVSFSDTIPYIQMNSY 558
QY 565 FTFOKCYCEILVYNEFGTGLNTCGSSNVSTTNMCAFTQGIQTEKTCGATSRFTNMF 624
DB 559 ISVYKGEVYLS-ITGLD---REDLHCDIDETCHP-QKSEALRELDEVMNAKYLIDF 612
QY 625 LILVSIPLAVILGIV--FKIR 645
DB 613 IVLGIFISLKLIAFVLKIR 635

RESULT 12

C86441
probable ABC transporter (imported) - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C86441
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federjpiel, N.A.; Kaul, S.; White, O.; A. Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewai ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; K. C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marz Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tai ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: C86441
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-646 <STO>
A:Cross-references: GB:AE005172; NID:g11136734; PIDN:AAG31315.1; GSPDB:GNO0141
C:Genetics:
A:Map position: 1
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1;

Query Match 17.7%; Score 589.5; DB 2; Length 646;
Best Local Similarity 29.9%; Pred. No. 2.8e-36;
Matches 183; Conservative 112; Mismatches 246; Indels 71; Gaps 22;

QY 10 GGSN--GLQVNRGSOS-SLEGAPATAPEPSLIGLTHASYSV-----HVRPMMWDTISCR 61
DB 14 GGVAVQGLPDMSPDQSSVLAFFPTISQ--GLQSMYPTTLKEVYKVK-TEQTSQCM 69
QY 62 QQW---TROIILKDVSLYVESGOIMCIGSSSGKTTLLDAMSGRLGRAGTFLGEVYNGR 118
DB 70 GSWKSKRTLLNGITGAVGCEFLAMIGPSGSGKTTLLSLALGSLR--TFSGVMYNG- 126
QY 119 ALRREQQDQF---STVLOSDTLLSLVRETLHTTALLAIRGNPGSFQK-----KVE 169
DB 127 ---QPSGCIKRTGTVAADVLYPHLTVWETLFFALLRL-----PSSLRDBKAEHVD 178
QY 170 AVMAELSLSHVADRILGNYSLGISTGERRRVSIAQLOLDPKVMLEDEPTTGLDCTAN 229
DB 179 RVIAELGLNCTNSMIGPLRFGISGGEKRVISIGQEMLNPILLDEPTSGIDSTYAH 238
QY 230 QIVLVELARRNRIVLTIHQPSSELPQLDKAIALISFGLICGTPAEMIDF 289
DB 239 RIYTTIKRLASGGRTVTTIHQPSSRITVHFDKAVLLSESPITYYGAASSAVEYFSSLG 298

```

OY 290 PCPEHNPEDFYMDLTS---VDIOSKEREIETSKRVOMIESAK-----KSAICHTL 339
DB 299 STSLTNPADLLDLNIPDPTOKETSEDOAKTVKTELVSAYEKNISTKLKAEICNAES 358
OY 340 KNIEEMK-HLKLTPVPEPKTKDSPGVESKLGVLRLVRYBNL---VKNLATVITLQML 395
DB 359 HSYETKRAANKLSEOMCT---WYQFVLQKRGVREKREPSFNKLRLEVOYISAEFL 414
OY 396 MGLLEFLFVLRNSVLKGAIDRVGLLYOFVGAATPYGMNAVLEPVLRAVSDOESOD 455
DB 415 GG--LLMW-----HTPKSHIQDRTALLFFSVMGVPLNNAVTFPOEKRMILKERS 466
OY 456 GLYOKOMMLAYALHVLFPVSVAATMISVVCYWTGLGHPVAREGYSALLAPHLIGEF 515
DB 467 GMYRLSYFARNGDLPLELALPTAEVFIYMMGSLKPDPTFELSLVLYSVLVAOG 526
OY 516 LTVLLGIVONPNTVNSVVALSTAGVNGSGLRNIOEMIPKRI-ISTYTPKXGSEI 574
DB 527 LGLAFGLALINIKATATLAVTLVFLIAGTY--VOQIP-PRIVMLKYSYSTYCYKL 582
OY 575 LVNVEFYGLNET 586
DB 583 LL-----GLOYT 589

```

RESULT 13

```

Probable ABC transporter [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 02-Feb-2001 #sequence, revision 02-Feb-2001 #text, change 16-Feb-2001
C:Accession: G84791
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; Vanhaken, S.E.; Unayam, L.; Tallon, L.;
Nuss, D.; Mierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: G84791
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-755 <STO>
A:Cross-references: GB:AE002093; NID:g4056489; PIDN:AAC98055.1; GSPDB:GN00139
A:Genetics:
A:Gene: At2g37360
A:Map position: 2
C:Superfamily: Arabidopsids thaliana probable ATP-binding cassette protein F12L6.1; ATP-I

```

```

Query Match 17.68; Score 584; DB 2; Length 755;
Best Local Similarity 27.28; Pred. No. 9e-36;
Matches 173; Conservative 128; Mismatches 253; Indels 82; Gaps 19;

```

```

OY 21 SOSLEGAPAT--APEPHSIGILHASYSGHRVPMMDITSCROQM-----TROILKDV 72
DB 79 SFSMASAPASISSSSPVLSTFDLTYSVKIQ-KKPNPLACCRSGNDSSVNTKILNGI 137
OY 73 SLVESGOICILGSSGGTTLTLDAMSGRIGRAGFLGAYVNGRALRREOFODCSYV 132
DB 138 SGEAREGEEMAIVAGASGSKSTLIDALANRIAR-DSLRSITLNGEVLSESMKVISAYV 196
OY 133 LQSDTLISSTVRETIHYTALLAIRGNPSFOKK-----VEAVMAELSLSHVADRILGN 187
DB 197 MODDLPEMLTVEETLWFSAEFRL---PRSLSKKKKARVVOALIDOLIGRSNAKFTVIGD 252
OY 188 YSLGISTGERRRVSIAQLODPPKVMLEPDETTGLDCTANQIVLVLELARNRNIVVL 247
DB 253 EGRGVSGGERRRVSINDIHPDILFLDEPTSGDSTSAVAVIKVQRIAGSGSLIVM 312
OY 248 THQPNSELPQDLKTAIISFGEILFCGPAEMLDENDCGYPCPEHNPEDFYMDLTSV 307
DB 313 SIHQPSYRIMGLDQDLFLSKNGTVYSGSPTHLPQFFSEKHPIDPENENKTEALDIL-- 370
OY 308 DTOSKEREIETSKRVOMIE-----SAYKSAICHTLKNIEEMKHLKLP----- 352

```

```

DB 371 ----RELEYSTEGTKPLVEFHKOWAKOABSYNNN---NKRNTVSSLKEATIASISGRK 423
OY 353 -----WVP-EKTKDSPGVESKLGVLRLVRYBNLVRNKLAVITRLQMLINGLF 399
DB 424 LVSCATNNNSNLTSPQTRANP-FWTEMTYIGKRALNSRQDELLGMKRGAVAVGII 482
OY 400 LLEFVLVRNSVLKGAIDRVGLLYOFVGAATPYGMNAVLEPVLRAVSDOESODLQY 459
DB 483 LATFTML-DNSPRGA-QERLG-PFAFAMSTFTYCAELPVLQEKYIMREAYANR 539
OY 460 KWOMMLAYALHVLFPVSVAATMISVVCYWTGLGHPVAREGYSALLAPHLIGEFYLV 519
DB 540 RSSYVLSOSIISIPALIVLISAFATFMVAVGLDGCANGGFEFFYTLISAFMAGSFRVF 599
OY 520 LIGIVONPNTVNSVVALSTAGVNGSGLRNIOEMIPKRIISTYTPKXGSEIIV 576
DB 600 LSGVI--PNWMLGTTVALL--AYFLFGGFPISRDRIPIVYMLFMHYSIVRYPEGVL 655
OY 577 VNEFYGLNETGSSNSVTTNPMCAFQGIQIETK 612
DB 656 QNEF-----QNPTRCFARGVQLDINS 676

```

RESULT 14

```

Hypothetical protein F5D21.6 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 02-Mar-2001 #sequence, revision 02-Mar-2001 #text, change 23-Mar-2001
C:Accession: D96533
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Al-
ansen, N.F.; Hughes, B.; Hultzer, L.
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim
C.A.; Li, J.H.; Lin, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maltl, R.; Marzi
Rizzo, M.; Rooney, T.; Rowley, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tall
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tall
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D96533
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-687 <STO>
A:Cross-references: GB:AE005173; NID:g10092349; PIDN:AA612758.1; GSPDB:GN00141
A:Genetics:
A:Gene: F5D21.6
A:Map position: 1
C:Superfamily: Arabidopsids thaliana probable ATP-binding cassette protein F12L6.1; A

```

```

Query Match 17.58; Score 583.5; DB 2; Length 687;
Best Local Similarity 28.08; Pred. No. 8.6e-36;
Matches 167; Conservative 117; Mismatches 239; Indels 73; Gaps 16;

```

```

OY 33 PEPHSIGILHASYSGHRVPMMDITSCROQM-----TROILKVSILYESGOICILGSS 88
DB 13 PPAEIG--RGALV-----WEDLVIVIPNFGSGTRRLDGLNGHAEGRIAMIMGCS 64
OY 89 GSKRTLLDAMSGRLGAGTFLGAYVNGRALRREOFODCSYVLOSDTLLSSTVRETL 148
DB 65 GSKSTLLDIAAGRLARNVIMTGNLLNGKARLD--YGLVAVVQEDILMGTLVRETI 122
OY 149 HTALLAIRGNPSFOKK-----VEAVMAELSLSHVADRILGNYSLGISTGERRVSI 203
DB 123 TYSAILRL-----SSDLTKREVNDIVGTTIETGLQCADRIVGNMHSRGVSGGERRRVSI 178
OY 204 AAQLDPPKVMLEPDETTGLDCTANQIVLVLELAR--NRNRYVLTIHQPSLELPQDLK 262
DB 179 ALIELRNPQILFLDEPTSGDLSAFVYQALRNIRNARDGGRVYSSIHQPSSEVPALFED 238
OY 263 IALLSFGEILFCGPAEMLDENDCGYPCPEHNPEDFYMDLTSVDTOSKEREIETSKRV 322
DB 239 LFLLSGGEYVYFGESKRAVEAFEAAGFPCPKRKNPSDHLRCINSDFIVATYATLKSGRI 298

```

```

OY 323 -----OMISAKKSAICHKTKLNIRMKHLTPMPVPEKTRDSP 362
DB 299 RETPATSDPLMNLATSEIKARLVEN-YRASYAKSARSRIELASIGHHMEVRKSEA 357
OY 363 GVFSKLGVLRRVYRNLRNKLAVITRLQNLGLLFFVLNRSNVLKGAIDRVGL 422
DB 358 TWFEQMLTKTRKSRVNNCRD-----IGYWSRIVYIYVSCVGTIFYDVGH 404
OY 423 LYQVVGATP-----YTQMLNANVL--PPVL--RAYDSQSDGLYOKQMOMLALAHVL 472
DB 405 SYTSILARVSCGGITFQFMFMSIGSGFPSTEEKVYKENSIGYGVSYIISNVSSF 464
OY 473 PFSVYATMIFSSVCYWTGLCHPEVARGYSALLAPHLIGELTTLVLGIVONPNIYNS 532
DB 465 PFLVALILIGISLIYNNMVKFRPGVSHMAFCLNIFSVSIESLMVYASLV--PNFLMG 522
OY 533 VVALLSIAG-VIVGSGFLRNIOEMPIPF--KIISYFTFOKYCSEILVNEFYGLNF 585
DB 523 LITGAGITGITIMTSGEFLRLPDLPKVFMRYPIFSMYSGMALQAVKNDFLGIEF 578

```

RESULT 15

JC7777

ATP binding cassette transporter, subclass G4 - human

C:Species: Homo sapiens (man)

C>Date: 02-Apr-2002 #sequence_revision 02-Apr-2002 #text_change 02-Apr-2002

C:Accession: JC7777

R:Engel, T.; Lorkowski, S.; Lueken, A.; Rust, S.; Schlueter, B.; Berger, G.; Cullen, P.; Blochem. Biophys. Res. Commun. 288, 483-488, 2001

A:Title: The human ABCG4 gene is regulated by oxysterols and retinoids in monocyte-deriv

A:Reference number: JC7777

A:Contents: Macrophages

A:Accession: JC7777

A:Molecule type: mRNA

A:Residues: 1-646 <ENG>

A:Cross-references: GB:AJ308237

C:Comment: This protein, a novel member of the ATP binding cassette(ABC) half-size trans

C:Genetics:

A:Gene: abcg4

A:Map position: 11q23.3

C:Keywords: macrophage

Query Match

Best Local Similarity 17.4%; Score 578.5; DB 2; Length 646;

Matches 171; Conservative 126; Mismatches 274; Indels 59; Gaps 14;

```

OY 33 PERHSGILHASYSHRVRPMWDITSCROQNTROLKDVSLYVSGQIMCITIGSGSGK 92
DB 54 PRRSANDIEFVELSYSVREGPCW-----RKRGYKTLKLSGKFCRRLIGIMPSGAGK 108
OY 93 TLLDAMSGRLGACFTIGEYVYNGRALRRBOFOCFYVLOSOTLLSLYRETLHYTA 152
DB 109 STEFNHILAGY-RESGMRQOILVNGRPRELRTFRKMSCTIMQDMILPHLTIVLEAMNVA 166
OY 153 LLAIRGNPGSFQKVEAV-----MAELSLSHVADRLIGNYSLGISGERRRVSIAAQ 206
DB 167 NLKLEKO-----EVKKELVTELTALGLMSCHSHTAL-----LSGGOKRLAIALE 214
OY 207 LLQDPKVLFDPEPTGLDQMTANOIVLVLLIARRNRIVLTIHQPSSELPQEDKIAVL 266
DB 215 LVNPNPVAFEDPSIGLDSASCFQYVSLMKSLAOGFTICTIHQPSAKLFEMFDKLYIL 274
OY 267 SEGELIFCGTAPAEIMDFENDCGYPCPEHSNPDFTMDLTSVDYOSKEREIETSKRVOMIE 326
DB 275 SOGOCIFKGVYTNLPIYKLGJLRCPYHNPAFLIEVASG-----EYGDLPMLF 325
OY 337 SAYKKSALCHTKLNIRMKHLKTLPMVPFKTKDSP-----GVFSKLGVLRLRRVTRN 378
DB 326 RAV-ONGICAMAEKSSPEKNEVPAPCPPEVPDIESHTFATSTLTQFCILFRKTFPLS 384
OY 379 LVKRLAVITRLQNLGLLFFVLNRSNVLKGAIDRVGLYQVVGATPYTGMLNA 438

```

```

DB 385 ILRDTVLHLRFNSHVIVGVLIGLILHIGDDAK--VENNTGCLFSSMLFLHFAIMPT 442
OY 439 VNLFPVLRVSDQSDGLYOKQMOMLALAHVL-PFSVATMTIFSSVCYWTGLCHPEVAR 498
DB 443 VLTFPLMAVFMREHLNMYSLKAYLAKTMDADVPQVVCPRVYCSIVYMTQGPATSTR 502
OY 499 FGYSALLAPHLIGELTTLVLGIVONPNIYNSVALLSIAGLVGSGFLRNIOEMPI 558
DB 503 FLFESALATATATVAQSLGL-LIGAAASNSIQVATFVGVPYAIPLVLFSGFVSFKTIPTY 561
OY 559 KTIISYFTFOKYCSEILVNEFYGL--NFTCGSSNYSVTNTMCAFTQIOPIETKCPG 615
DB 562 LQMSSTLYSYRIGFEGYIL-ITYMERGDLTC-----LEERCPRREP-QSILRALDV 611
OY 616 ATSRFTMNLILYSFIPALVILGIYVEKIR 645
DB 612 EDKILYMDFLVLGIFLALRLRLAYLVLYLR 641

```

Search completed: June 11, 2003, 09:13:51
 Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 09:08:12 ; Search time 39 Seconds

(Without alignments)
3439.402 Million cell updates/sec

Title: US-09-989-981A-6

Sequence: 1 MGDLSLPFGSGMKGLQVNRG.....PALVILGIYVFKIRDLISR 651

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: SP:REMBL_21.*
2: SP:archaea.*
3: SP:bacteria.*
4: SP:fungi.*
5: SP:human.*
6: SP:invertebrate.*
7: SP:mammal.*
8: SP:mhc.*
9: SP:organelle.*
10: SP:phage.*
11: SP:plant.*
12: SP:rodent.*
13: SP:virus.*
14: SP:vertebrate.*
15: SP:unclassified.*
16: SP:viirus.*
17: SP:bacteriap.*
17: SP:archaeap.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	704	21.2	673	11	Q8R543
2	680.5	20.5	655	4	Q867A8
3	672.5	20.2	655	4	Q961D6
4	662	19.9	801	5	Q8T691
5	660	19.8	657	11	Q9R004
6	649.5	19.5	725	10	Q9M3D6
7	628	18.9	691	10	Q8RW19
8	622	18.7	668	10	Q9ARU4
9	618.5	18.6	672	10	Q9ALR2
10	617	18.6	727	10	Q9FNB5
11	615	18.5	692	5	P91892
12	614.5	18.5	703	10	Q8RXN0
13	614	18.5	594	10	Q8RLC3
14	614	18.5	720	10	Q9M2V7
15	610.5	18.4	725	10	Q9Z035
16	610.5	18.4	725	10	Q9ASR9

17	610	18.3	679	5	Q9BH97	Q9bh97 ceratitlis c
18	608	18.3	708	10	Q9M2V5	Q9m2v5 arabidopsis
19	602.5	18.1	654	10	Q9LIW2	Q9liw2 oryza sativ
20	600.5	18.1	670	5	Q77423	Q77423 bacterocera
21	600	18.0	602	5	Q9VC15	Q9vc15 drosochila
22	597	17.9	590	10	Q9MAH4	Q9mah4 arabidopsis
23	595.5	17.9	658	5	Q16574	Q16574 caenorhabdi
24	595.5	17.9	687	5	Q94960	Q94960 drosophila
25	595.5	17.9	785	4	Q96176	Q96176 homo sapien
26	592	17.8	740	5	P90746	P90746 caenorhabdi
27	591.5	17.8	740	10	Q80946	Q80946 arabidopsis
28	589.5	17.7	646	10	Q966R7	Q966r7 arabidopsis
29	588.5	17.7	648	10	Q966W5	Q966w5 arabidopsis
30	587.5	17.7	627	11	Q91MA3	Q91ma3 mus musculu
31	587.5	17.7	627	11	Q91MA3	Q91ma3 mus musculu
32	584	17.6	755	10	Q9ZUT0	Q9zut0 arabidopsis
33	583.5	17.5	687	10	Q9C8K2	Q9c8k2 arabidopsis
34	581.5	17.5	687	5	Q9NH94	Q9nh94 bombyx mori
35	580.5	17.5	669	5	Q8WRE2	Q8wre2 tribolium c
36	578	17.4	705	10	Q9LIW1	Q9liw1 oryza sativ
37	575.5	17.3	649	10	Q9SIT6	Q9sit6 arabidopsis
38	575.5	17.3	666	11	Q9EPG9	Q9epg9 rattus norv
39	573.5	17.2	669	5	Q8WRE1	Q8wre1 tribolium c
40	572	17.2	739	10	Q9LFC8	Q9lfc8 arabidopsis
41	571.5	17.2	609	10	Q9C8W6	Q9c8w6 arabidopsis
42	571.5	17.2	662	10	Q949Y4	Q949y4 drosophila
43	571	17.2	609	5	Q9YON4	Q9yon4 dictyostell
44	567	17.0	626	5	Q8T684	Q8t684 dictyostell
45	565	17.0	1528	5	Q8T677	Q8t677 dictyostell

ALIGNMENTS

RESULT 1
ID Q8R543 PRELIMINARY: PRT: 673 AA.
AC Q8R543;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Steroidin 2.
GN ABCG8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RA Lu K., Zhou Y., Lee M.-H., Patel S.B.;

RT "Molecular cloning, genomic structure and characterization of novel
mouse head-to-head tandem ABC transporters";
RT Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF351811; AAL82898.1;
DR EMBL: AF351799; AAL82898.1; JOINED.
DR EMBL: AF351800; AAL82898.1; JOINED.
DR EMBL: AF351801; AAL82898.1; JOINED.
DR EMBL: AF351802; AAL82898.1; JOINED.
DR EMBL: AF351803; AAL82898.1; JOINED.
DR EMBL: AF351804; AAL82898.1; JOINED.
DR EMBL: AF351805; AAL82898.1; JOINED.
DR EMBL: AF351807; AAL82898.1; JOINED.
DR EMBL: AF351808; AAL82898.1; JOINED.
DR EMBL: AF351809; AAL82898.1; JOINED.
DR EMBL: AF351810; AAL82898.1; JOINED.
SQ SEQUENCE 673 AA: 76008 MW: FA08340445DE259C CRC64;

Query Match 21.2%; Score 704; DB 11; Length 673;
Best Local Similarity 28.7%; Pred. No. 5.8e-43;
Matches 195; Conservative 130; Mismatches 261; Indels 94; Gaps 17;


```

Db 14 GVLADASGGLDSDL-----ESSESNSTLYTSSGSGNLTVELDLYOVDIASQVWFEOI 69
Oy 58 TSCROQWTRQI-----LKDVSLYESGOIMCIGSSGSKTLLDMSGRILRAGTF 109
Db 70 AQFKIPMRSHSSODCELGIRLISFKVNSGOMALITGSSGCRASLADVIYTR-CHGCKM 128
Oy 110 -LGEVYVNGRALRBOQDFCSYVLQSDTLASLTLYRETLHYTALLAI-RGNGPSFQK 167
Db 129 KSGQIMINGQSTPOLVKRCVAHVROHQLLPNLVRETLAIQMLRPTFSQQRDR 188
Oy 168 VAVVAVALSLSHVADRLIGNSLGISTGERRRVISAQLODPKVMLEDEPTGLDCMT 227
Db 189 VEDVIAELRLRQCACATRGNTVYRGVSGERRRVSIGVLMNPGLILDEPTSGIDSET 248
Oy 228 ANQIVLYVELARRRRIIVLTIHOPRSELPOLFQIALISFGLJFCGTPAMLDPEFNC 287
Db 249 AHNLTVTLRLAKGRVLVLSLHOPRSDIFLELDVLTMTSTPIYLGAQOMVOYFYSI 308
Oy 288 GYCPESHNPDEYMDLTSVDTQSKERELETSKRVOMIESAYKKA-----ICHKTKN 341
Db 309 GHPCRYSNPADFYDLTSDIRSKEREVAAYEKQSLAFLLEKVGDFDLKMAEKE 368
Oy 342 IERMAHLKTLPRVPRKTKDS-----PGVFSKGLVLRVTRNLVNRKLAIVTRLLON 393
Db 369 LNTSTHTVSLTL---TODTDCGTAVLELPGMIEQSFSTLIRROISNDFRDLPTLIHGEA 424
Oy 394 LIMGLELFEVLVRVNSVLKGIODRVGLIYOFVATPYTGMLNVLNPEVLRASDDES 453
Db 425 CLMSLITIGFLYGHAKOL--SEMDTALFLPMIGALIPFNVLIDVYSSCHSERSMLYEL 482
Oy 454 QDGLYQKQOMLAVLHVLPESVAVATMIFSSVCYTLCILHPEVARFGFSALLAPHLIG 513
Db 483 EDGLTAGRYFFAKILGELPEHCAYVIYAMPIYMLNLRP-----VPELFL 529
Oy 514 EELTVILGIVONPPIVNSVALLS-----IAGVLYGS-----GFLRNIQEMDIPKRI 561
Db 530 LHELVMLVVFCCRNMLAASAMLPFMSSFCNALNSFTLGTAGFMINDMLIVAM 589
Oy 582 ISYTFQKYSSEILVVFNEFYGLNFT--CGSSNVSTYTPMCAFTGIGIETKCGATSR 619
Db 590 ISKLSFLMKRCSGLMOIOPNGHLTYTOIGNFTFSILGDM-----ISA 632
Oy 620 FTMNPLIYSPFIPALVILGI 639
Db 633 MDLNSHPLYAYI-LIYIGI 650

```

RESULT 2

0967A8 PRELIMINARY; PRT; 655 AA.

AC 0967A8; MEDLINE-21201983; PubMed-11306452;

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE ATP-binding cassette superfamily G (White) member 2.

GN ABCG2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_Taxid-9606;

RN NCBI_Taxid-9606;

RP SEQUENCE FROM N.A.

RT resistance/placenta-specific, ATP-binding cassette transporters as a

RT transporter of NB-506 and J-107088, topoisomerase I inhibitors with an

RT indolocarbazole structure.

RT Cancer Res. 61:2827-2832(2001).

DR EMBL; AB051855; BAB46933.1; -

DR InterPro; IPR003439; ABC_transportr.

```

DR InterPro; IPR003880; Pantane_attach.
DR InterPro; IPR003580; Protachykinin.
DR Pfam; PF00005; ABC_tran.1.
DR ProDom; PD000006; ABC_transportr.1.
DR SMART; SM00203; Trk.1.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_1.
KW ATP-binding.
SQ SEQUENCE 655 AA; 72314 MW; ABA6F6B96034C5A8 CRC64;

Query Match      20.5%; Score 680.5; DB 4; Length 655;
Best Local Similarity 29.2%; Pred. No. 2.9e-41;
Matches 182; Conservative 137; Mismatches 250; Indels 55; Gaps 18;

Oy 21 SSSLEGAPATP---EPISGLILNAYSVSVRVPMMDITSCROQWTRQIKDYLVE 77
Db 13 SGNNGNPFATASNDLKAFTBEGAVLSFNHICRYALKSGFLPCRPAVEKELISNIGTK 72
Oy 78 SCQIMCIGSSGSGKTTLLDAMSGRLGRAGTFLEGYVNGRALRBOQDFCSYVLQSDPT 137
Db 73 PG-LNALIGPTGGKSSLLDVLAAKRDPSG-LSGDVLING-APRANFRCNSGYVQDDV 129
Oy 138 LLSLTVRRLTYTALLAIRGNPG-SFOKKEAVMAELSLSHVADRLIGNSLGISTG 196
Db 130 VMGTLTVRENLQFSALRLATTMTNHEKERNIRNYOELGLDKVADSKVGTQFIRGVS 189
Oy 197 EERRVISAQLODPKVMLEDEPTGLDCMTANQIVLYVELARRRRIIVLTIHOPRSEL 256
Db 190 EKKRISIGMELTDSILFLEDEPTGLDSTANAVALLKRNKSGKRTTFSIHOPRST 249
Oy 257 FOLFPAKIALISFGLIFCGTPAEMLEFNDGCPCEPHSNPPDFYMDLTSVTO-----SK 312
Db 250 FKLFDLSLLTAGRLMFGHPAQEALGYFESAGYHCEAVNNPADFLLIDNGSTVALNR 309
Oy 313 ERE-----IENSKR-----VOMIESAYKKAICHT-----LKNIERKHLKTLPMVPF 356
Db 310 EEDFKRATLIEPSKODKPLIEKLAELIYVNSPFYKTEKAKELHOLSGGEKKKTKVREISY 369
Oy 357 RTKDSPGVFSKGLVLRVTRNLVNRKLAIVTRLLONLIMGFL--LEFVLVRVNSVLKGI 414
Db 370 TT-----SCHOLRWKSKSFRLNLPDASIAQIIVYVVLGLVIAITIFGLANDST--- 421
Oy 415 AIODRVGLLYOFVATPYTGMLNVLNPEVLRASDDESODGLYQKQOMLAVL-HVLP 473
Db 422 GIONRAGVLF-FLTNGCFSSVSAVELVEVEKELIHYISGYVSSYPLGKLSDLLP 480
Oy 474 FSVVATMIFSSVCYTLCILHPEVARFGFSALLAPHLIGELTVILGIVONPPIVNSV 533
Db 481 MEMLPSTILFTCIYVMLGLKPKADAFYVMFTLM--WVAYSASSMALAIAAGOSVSA 537
Oy 534 VALLSIAGV--LVGSGFLRNIQEMDIPKRIISYTFQKYSSEILVVFNEFYGLNFTCGSSN 591
Db 538 TLMTICFVPMAMIFSGLVNLTITASLMSLMOYFSIPRYGFTALQHNELGONFCPG--- 594
Oy 592 VSVTTNPMCAFTGIGIETKCPG 615
Db 595 INATGNMPCNVA-----TCTG 610

```

RESULT 3

096LD6 PRELIMINARY; PRT; 655 AA.

AC 096LD6; MEDLINE-21201983; PubMed-11306452;

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE ABC transporter ABCG2.

GN ABCG2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_Taxid-9606;

RN NCBI_Taxid-9606;

RP SEQUENCE FROM N.A.

RA Schuetz J.D., Wall A.M., Sampath J., Sorrentino B., Du G.,
 RT "The Human ABC Transporter, ABCG2, Transports Hoechst 33342 and
 RL Requires an Intact Walker A Motif.",
 DR Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY017168; MG32982.1;
 DR InterPro: IPR003439; ABC_transport.
 DR InterPro: IPR003880; Ppantne_attach.
 DR Pfam: PF00005; ABC_tran. 1.
 DR ProDom: PD000006; ABC_tran. 1.
 DR PROSITE: PS00012; PHOSPHOPANTHEINE; UNKNOWN.1.
 SO SEQUENCE 655 AA; 72288 MW; B3B5DC02C095C4A8 CRC64;

Query Match 20.2%; Score 672.5; DB 4; Length 655;
 Best Local Similarity 29.0%; Pred. No. 1.1e-40;
 Matches 181; Conservative 137; Mismatches 251; Indels 55; Gaps 18;

QY 21 SSSLEGAPATAP---EPHSLGILHASYSVSHRPRPMDITSCROOITROILKDVSLYE 77
 DB 13 SGGNTNGFPATASNDLKAFTGEGAVLSFNHCYRVKLSGFLPKRKEVEKILSNINGIMK 72
 QY 78 SGOIMCIGSSGSKTLLDAMSGRLGRAGTLEGYYVNGRALRRDQFDCSYVLQSDPT 137
 DB 73 PG-LNALGPPTGGKSLDLVLAARKDPG-LSGDVLING-APRANFKCNSGYVQDDV 129
 QY 138 LLSLVRRETLHYTALLAIRGNPG-SFOKVEAVMAELSHVADRILGNISLGISTG 196
 DB 130 VNGTLVRNENLQFSALRLATTTWTHNEKERINRYOELGLDKVADSKGTQFIRVSGG 189
 QY 197 ERRRSIAQQLQDPKVMLEPTTGGLDPMANOIVLLVELARRNRYVLTTHDRSEL 256
 DB 190 EKKRSISGIMELTIDPSILFLEDEPTGLOSTANAVLLKRSKOGRTIIFSIHOPRST 249
 QY 257 FOLFKAIRLSCGLIFCSTPAEMLDFRDCGCPCEPHSNPPEFYMDLTSVDTO---SK 312
 DB 250 FKLFDSTLTLASGRMLFHPGPAQELGYESAGYHCEAVNPNADFLDIINGDSTAVALNR 309
 QY 313 ERE-----IETSKR-----VOMIESAYKKSACHT-----LKNIEKRLKLTLPMP 356
 DB 310 EEDFATATEIIEPSKODKDLIEKLAITYNNSFYETKAKELHQLSGEKKKTYVEKEISY 369
 QY 357 KTKSPGVFSKGLVLRNRYLRNKLAVITRLQNLIMGLF-LFVFLRVRSNVLKG 414
 DB 370 TT---SECHOLRWKSRKFLNLCNPQASIAQIIIVYVGLVIGAITFGKNDST--- 421
 QY 415 AIODRVGLLQFVGATPTTGMNANVLFVPLRAVSDOSSDGLYQKQWMLALAYL-HVLP 473
 DB 422 GIQNAQVLF-FLTTNOCFSSVASVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLP 480
 QY 474 FSVVATMIFSSVCTWTGLHPEVARFGYSAALAPHLIGELFLVLVLIQONPRTYNSV 533
 DB 481 MRMLPSTIEIYVFMGLKAKADAFVYMMFTL---WVASASSMALAIAGGSVSA 537
 QY 534 VALLISAGV-LVSGGLRNIOEMPIFKIISYTFPOKCEILVWVEFVGLNFTCGSSN 591
 DB 538 TLMLTICFVEMIFSGLLVNLJTJIASWLSMLOYSIRYGTALQHNHEFLQONPCG--- 594
 QY 592 VSVTTNMCATGQIGTIEKTCPG 615
 DB 595 LNAIGNPCNVA-----TCTG 610

RESULT 4

ID 08T691 PRELIMINARY; PRT; 801 AA.
 AC 08T691;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DE ABC transporter Abcgl. 21, Last annotation update)
 GN ABCG1.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
 OX NCBI_Taxid=44689;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AX4;
 RA Anjard C., Loomis W.F.;
 RT "Evolution of the ABC transporters of Dictyostelium";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF482380; AAL91485.1;
 SO SEQUENCE 801 AA; 90052 MW; CCCAF0036CB195A3 CRC64;

Query Match 19.9%; Score 662; DB 5; Length 801;
 Best Local Similarity 27.4%; Pred. No. 8.3e-40;
 Matches 185; Conservative 134; Mismatches 246; Indels 110; Gaps 16;

QY 61 RQOITROILKDVSLYESSGOIMCIGSSGSKTLLDAMSGRLGRAGTLEGYYVNGRAL 120
 DB 131 KKKISQILNINIGHIESGTIFAIMGPSGAKTLLDIHLRLININGS--GFWYLNKNS 188
 QY 121 RREOFDCSYVLQSDPTLLSLVRETLHYTALLAIRGNPGSFO-KYEVMAELSLH 179
 DB 189 DFNIFKLGQYVQSDSLMPSLVRRETLNFAQLKMPROVPLAKELQRODITIDENGLNR 248
 QY 180 VADRLIG--NYSLGISTGERRRVSIAQQLQDPKVMLEPTTGGLDCTANOIVVLYVE 237
 DB 249 CADTLVGTADNKRIGISGERRRVTISIELLGPVYLLDEPTSGLDASTFYVSAALK 308
 QY 238 LARRNRYVLTTHDRSELFOLEDKAITSFGEELFCSTPAEMLDFRDCGCPCEPHSNP 297
 DB 309 LAKSGRTTCTTHOPSNITDMDNLGLDGTIYGRANKALEFANGYHCSKTNP 368
 QY 298 PDEFYMDL-TSYDTOS----- 311
 DB 369 ADFPDLITQYVEDADSDDDYNDDEEETGGGGGGGAGIEDIGISPTMNGSAV 428
 QY 312 ---KERIE-----TSKRVQMIESAYKKS---AICHKTLKN 341
 DB 429 DNKNNEELKQOQOQOQOQOQOQSTGRRARRIKKLTREEMVILKKEYPNSOGILRVETLDN 488
 QY 342 IER-----MKHLKTLPMPPFKTKDPGVFSKGLVLRNRYLRNKLAVITRLQNLIM 396
 DB 489 ISKERTRPKTEKT-----RGNPFLTQESLLGHEVYNAKHPMAFKVNLQALFQ 539
 QY 397 GLFLFVLRVRSNVLKAGAIODRVGLLQFVGATPTTGMNANVLFVPLRAVSDOSSDG 456
 DB 540 G--LLCGIYVYQOLGIGQSSVOSRTVAFITMGVSFPVMSITHFPPVITFLDRASG 597
 QY 457 LYOKQWMLALAYLHVPSPVATMIFSSVCTWTGL--LHPEVARFGYSAALAPHLIGE 514
 DB 598 VYDLPFLFLAKSPMDACIAVLLPMTATIVYMMTORDRPYSAAPFRFLM---LVLA 654
 QY 515 FLTLVLGIVON---PNI-VNSVALSIACVLVSGGLRNIOEMPIFKIISYTFPOKY 570
 DB 655 SOTCLISGLVLISSVPNVQVGTAAVPLVLIIFLFSGFFIMNDVPGVLWVEPISFRY 714
 QY 571 CSEILVWVEFVGLNFTCGSSNVSVTNMCATGQIGTIEKTCGARSFRPMNLIYSF 630
 DB 715 MIEAANVNAFQDVFTCTDSO---KIGGVCYQYGNVYE-NMGYDIDHFRNRWIMLVLY 770
 QY 631 IPALVILGIVVEKIR 645
 DB 771 IIGRRVLTFLVLKIK 785

RESULT 5

ID 09R004 PRELIMINARY; PRT; 657 AA.
 AC 09R004;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DE Breast cancer resistance protein 1.
 GN ABCG2 OR BCRP1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-FVB: TISSUE=LIVER;
 RX MEDLINE=99413474; PubMed=10485464;
 RA Allen J.D., Brinkhuis R.F., Wijnholds J., Schinkel A.H.;
 RT "The mouse Bcrp/Mxr/Abcp gene: amplification and overexpression in
 cell lines selected for resistance to topotecan, mitoxantrone, or
 doxorubicin";
 RL Cancer Res. 59:4237-4241(1999).
 DR EMBL: AF140218; AAD54216.1; -.
 DR MGI: MGI:1347061; Abcg2.
 DR InterPro: IPR003439; ABC_transportr.
 DR InterPro: IPR003880; Ppantne_attach.
 DR Pfam: PF00005; ABC_tran.1.
 DR ProDom: PD000006; ABC_transportr.1.
 DR PROSITE: PS00012; PHOSPHOTRANSFERASE;
 SQ SEQUENCE 657 AA; 73021 MW; 207B70BC272CC0D5 CRC64;

Query Match 19.88; Score 660; DB 11; Length 657;
 Best Local Similarity 28.08; Pred. No. 9e-40;
 Matches 181; Conservative 135; Mismatches 242; Indels 88; Gaps 19;

QY 13 MGLVNRGSSSLGAPATAPRPHSGLHASYSVSHRVPMWDITSCROQWTOILKDV 72
 Db 12 MSQNNNGQLPRMNRARVTLAEGDVLSEFHITTYRV--KVKSGFLV--RKVEKEILSDI 66
 QY 73 SLVYESQIMCITLSSSGSKTTLTLDAMSGRLGRGTGLGEVYVNGRALRQPODCSYV 132
 Db 67 NGIKPKP-LNALITLPTGGKSSSLDVLAAKRDPRG-LSGDVLNG-APPAHFKCCSGYV 123
 QY 133 LQSDTLSSLVRETLHYVTALLAIRGNPCSFQ-----KKEAVMAELSHVADRILGN 187
 Db 124 VQDVVVGTLVRRNIQPSALRL-----PTMKHKEKERNITIKELGEKVDKSVGT 179
 QY 188 YSLGISTGERRRVSIQAQLQDPKVMLEPTTGLDQMTANOIVLVLELARRNRIVVL 247
 Db 180 QFINGISGGERKRSISMEILTPDSILFLDEPTGLDSTFANAVILLKLRMSKGRITIF 239
 QY 248 TIHPRSELPFLPKIALISFGLFCGTAPRMDLDFNDCYPCPEHSNPFDFMDLTSV 307
 Db 240 SIHPKRSIFKFLPSLTLASGLKVFHPQAKALEYRASAGYHCEPYNPAPDFLDVING 299
 QY 308 DTOS-----KERETSKR-----VOMISAKKSAICKTKLNIERRMKHLTLP 352
 Db 300 DSSAVMLNRBQDNEANKTEPEKGVLENSEFIYNSALVGETKAELDQ----- 352
 QY 353 MWPKTKDSPGVESKLV-----LIRVTRNLVRRKLAIVITRLQNL 394
 Db 353 -----PGAQKKKSTSAFKEPEVYTSFCHQLRMIAKRSFKMLGNPQASVQALLVTV 403
 QY 395 IMGLFL--LPEVLVRSNVLKGATODRVGLLYOVGAPTYTGMLNANVLFPVLAASDOE 452
 Db 404 ILGLITGLITFDLYDA---AGMONRAGVLE-PLTNQCFSSSAVELLEVVERKLTIE 458
 QY 453 SQDLXYKQWMLALVAL-HVLPFSVAVMIFSSVCYVTLGLHPVARFGYSALLAPHL 511
 Db 459 YISGYRVSYSFEFGKVSMDLLPMRFLPSVIFTCLLYMLGKAKTVDAFFIMFLI---M 515
 QY 512 IGEFLTLVLYLQYONPVIWVSVALSLIAGV--LVSGGLRNIOEMDIPKRIISYFTFQK 569
 Db 516 VAYTASSMALAIATGOSVSVATLMTIAEFVFMILFGLLVNLTITIGPMLSMILQYFSIPR 575
 QY 570 YCSLIVNFEYGLNFTCGSSNVSVTNPMCAFTQGIQFIETKCPG 615
 Db 576 YGFTALQYNEFLGOEFCPG--FNVTNSTCVNSYAI-----CTG 612

RESULT 6
 ID 09M3D6 PRELIMINARY; PRT; 725 AA.
 AC 09M3D6;

DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE ABC transporter-like protein (Putative ABC transporter protein).
 GN T26112.10 OR AT3G55130.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RN SEQUENCE FROM N.A.
 RA Monfort A., Casacuberta E., Pulgomenec P., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X., Quetier F., Salanoubat M.;
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RN SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RA Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RN SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Goldsmith A.D.,
 RA Jiang P.X., Lee J.M., Onodera C.S., Quach H.H., Tang C., Tortum M.,
 RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamita A., Karlin-Neumann G.,
 RA Kawai J., Kim C., Koesema E., Lam B., Lin J., Meyers M.C., Miranda M.,
 RA Natsuka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
 RA Shinn P., Southwick A., Tracy S.E., Shinozaki K., Davis R.W.,
 RA Ecker J.R., Theologis A.;
 RT "Full length cDNA of gene T26112.10/AT3G55130 (GI:7019646).";
 RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RN SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Etgu P., Lee J.M.,
 RA Tortum M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,
 RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,
 RA Southwick A., Davis R.W., Ecker J.R., Theologis A.;
 RT "Arabidopsis Open Reading Frame (ORF) Clones";
 RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AL132954; CAB5747.1;
 DR EMBL: AY045932; AK76606.1;
 DR EMBL: AY079387; AL85118.1;
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transportr.
 DR Pfam: PF00005; ABC_tran.1.
 DR ProDom: PD000006; ABC_transportr.1.
 DR SMART: SM00382; AAA.1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN.1.
 DR ATP-binding.
 DR KW
 SQ SEQUENCE 725 AA; 80656 MW; 790C535A7929C16 CRC64;

Query Match 19.58; Score 649.5; DB 10; Length 725;
 Best Local Similarity 29.48; Pred. No. 6e-39;
 Matches 182; Conservative 124; Mismatches 246; Indels 67; Gaps 15;

QY 33 PEPSLGLHASYSVSHRVPMWDITSCROQWTOILKDVSLVYESQIMCITLSSSGSK 92
 Db 68 PVEYVLFNNLQYVITARR-----FGSRQNGKVTLLDDVSGASDSDIILAVLAGASAGK 123
 QY 93 TLLDAMSGRLGRGTGLGEVYVNG-RALRQPODCFSYVLSQDTLSSSLVYRETLHYT 151
 Db 124 STLDALAGRAVE-GSLRGSVTLNGEVLKSLRLKLVISAVYMDDLFLPLTYKETLMPA 182
 QY 152 ALTAIRRG-NPGSQKQVEAVMAELSHVADRILGYSLSGISTGERRRVSIQAQLQD 210
 Db 183 SEFLPRLSLSKSKMEVEALLDQLGRNANVTYDGEHGVSGGERRRVSIGIDIHD 242
 QY 211 PKVLFDEPTTGLDQMTANOIVLVLELARRNRIVTLTHQPSSELTQLFDKTAISFGE 270
 Db 243 PIVFLDEPTSGDSTAFANVYVYLKRIAGSGSVIWSIHQPSARIYELDLRLITLSRGK 302
 QY 271 LIFCGTAPAEMLDFNDCYPCPEHSNPFDFMDLTSVDTSKREIETS----- 319

Db 303 SVFNGSPASLPGFSDGRIPEKENISEFALDIV-----RELEGSNEGTRALVDEN 354
 QY 320 -----KRVOMIESAYK-----KSAICHKTL--KNIEEMKHLKLPMPFETKD 360
 Db 355 EKMOONKISLISQSPQTKNDODBSLSLKEAINASRGKLVSGSSSNPTSMETVSSYA 414
 QY 361 SPGVESKGLVLRVRYRNRLVKNLAVITRLQNLIMGLFLFEVLRYRSNVLGAIDRV 420
 Db 415 NPSLFEFF-ILAKRYKMNWIMPELVGTRIATVMTGC--TLATVYKMDLHTPPGA-QERL 471
 QY 421 GLIYQFVGATPYTGMLNAVLFPVLRAVSQESQDLDYOKOMMLAYALHVPFVATM 480
 Db 472 -TLFAFVPTMFCCDNDNPVPIQERYITLRETHNAYRTSSVYSHSLPQLAPSL 530
 QY 481 ISSVCYWTGLGHPVAREGYSFSAALLAPHLIGEFLTVLGLIYQNPVINS-VVALSTI 539
 Db 531 VESAITFMTVGLSGLEGEFVFCILYASFSGSSVPTISGV--PNTMLCYMVSITYL 588
 QY 540 AGVLGSGFLRMIOEMPIPKIISTFTQKYCEILVNEFYGLNFTCGSSNVSTTNPM 599
 Db 569 AYCLLSGTYVNRDRIPEYTWTFHYISILKYEAVLINEF-----DDPS 633
 QY 600 CAFTOGIOFIKTCGATS 618
 Db 634 RCFVRGVQVFDSTLLGVS 652

RESULT 7

Q8RM19 PRELIMINARY: PRT: 691 AA.
 AC Q8RM19;
 DT 01-JUN-2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Hypothetical.77.2 kDa protein.
 GN AFG21090.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 RX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carandini P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamlya A., Kawai J.,
 RA Kim C., Lin S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shino P., Yamada K., Shinozaki K.,
 RA Becker J., Theologis A., Davis R.W.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY093054; AAM13053.1;
 KW Hypothetical protein.
 SQ SEQUENCE 691 AA: 77219 MW: 6E473CC0B40D7E9 CRC64;

Query Match

Best Local Similarity 18.9%; Score 628; DB 10; Length 691;
 Matches 173; Conservative 123; Mismatches 225; Indels 94; Gaps 17;

QY 25 LEGAPATAPF-PHISGILHASYSHRRPMDITSCROO---TRQILKDVSLYVSG 79
 Db 3 LEGSSSGRQPLSKLEMSRGAVLA-----WEDLVVVPNSDPTRLRLQRLNGYAE 56
 QY 80 QIMCIISSGSGKTTLLDAMSGRLGRAGTFLGEYVNGRALRRQFDCSYVLQDPTL 139
 Db 57 RIMAIMPGSSGKSTLLDSLGRARNAVMTGNLLNGKARLD--VGLVAYVQEDVLL 114
 QY 140 SSLVRETLHTALAIRNPGSFQK-----VEAVMAEISHVADRILIGNSLQGIS 194
 Db 115 GTLVRETIYSALRL-----PSDMKREYSDIVEGTIIELGLDCCSDRYIGMHWARG 170
 QY 195 TGERRRSIALQLODPKVLDPDEPTGLDCMTANOIVLLBELARRNRIVLTIIHQPRS 254
 Db 171 GGRKRNVSTALETLRPOILFDEPTSGDLSASAFVQLRLNIARDGRIVISSVQPPS 230

QY 255 ELRQLPDKIALISNGELIFGCTPAEMLDFFNDGCPDPEHNSPDPFYMDLTSVDTOSKER 314
 Db 231 EVRALFDLFLSSGESVYGEAKSAVEFAESGFCPCPKRNPSDHLRLINSDFDTVA 290
 QY 315 EIETSKRVQ-----MIEAAYKSAICHKTLKNIEEMKHLKLPV 354
 Db 291 TLKSGRIQETPATSDPLMLATSVIKARLEN-YKRSKAKSARIRRELINIEGLEME 349
 QY 355 PFTKSPGVESKGLVLRVRYRNRLVKNLAVITRLQNLIMGLFLFEVLRYRSNVLKG 414
 Db 350 IRKSGEATW-MKQLRFLTRASFIMCRDQVYTRIIISYIVSI----- 392
 QY 415 AIQDRGGLLYQFVATPYTGMLNAVL-----FPVL---RAVSDOSQDS 456
 Db 393 ---SVGTIFVGVGS-YTSILARVSCGGEFTGFMTEMSIGFSPLEMKRYKRLSG 447
 QY 457 LYOKQOMMLAYALHVPFVSVAQDMIFSSCYTTLGHPVARNRGFSALLAPHLIGEFL 516
 Db 448 YGVGSVYILSNISSEFPFLVALSVITGITTYMLVFERPQFSHAFCLNIFPSVYEST 507
 QY 517 TLVLGIVQNPVINSVYALLSIAG-VLVGSEFLNIOEMPIPKI-----ISTFTQKY 570
 Db 508 MMYVASVY--PRLKGLITGNGLIGIIMATSGFFRLDLP---KIFRYPVSYSTYSGSW 562
 QY 571 CSEILVNEFYGLNF 585
 Db 563 AIOGGYKNDFLGLEF 577

RESULT 8

Q9ARU4 PRELIMINARY: PRT: 668 AA.
 AC Q9ARU4;
 DT 01-JUN-2001 (TReMBLrel. 17, Created)
 DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Putative ABC transporter.
 GN P0445D12.3.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ox
 RX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 KL clone: P0445D12.3;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

CC -1. SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL; AP003046; BAB40032.1;
 DR InterPro: IPR003593; AAA_Artpase.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transport; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Transport.

QY 55 WDTISC-----RQMTROIILKDVSLYVSGQIMCIISSGSGKTTLLDAMSGRLGRAGT- 108
 Db 58 WANITCALKNRKDDVARFLLSNASGEAKSGRLALAMPSSGKTTLLANVLAGOLVSPSL 117
 QY 109 -FLGEYVNGRALRRQFDCSYVLQDPTLLSLVRETLHTALAIRG--NPGSFQK 166
 Db 118 HLGSEFLXINRPISEGCYK--IAYVROEDLFEFQTLVRETS- 170

Query Match 18.7%; Score 622; DB 10; Length 668;
 Best Local Similarity 31.0%; Pred. No. 5.4e-37;
 Matches 171; Conservative 100; Mismatches 245; Indels 36; Gaps 12;

```

Oy 167 KVEAVMAELSLSHVADRLICNYSLGISGTGERRRVINAQLLODPKVALFDEPTTGDCM 226
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 176 YVNDLFLRLGLADNCADSIYDAKVRGISGGEKRLSLACELIASPSIIFADPEPTGDAF 235
Oy 227 TANOIVLVLLARNRKRVLTITIQPSSELFQLEDKTAISFGELIFCGTGA--EMLDFE 284
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 236 QAEVFMETRLQADGHTVICSIHQPSGVYKGFDDIVLSEGEVIYMG-PAKEPEPLXF 294
Oy 285 NDCGYPCPEHNSPDFYMDLTSVDOSKERIEITSKRYQ-MIESAKKSAI--CHKTKLN 341
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 295 ASLGIHCPDHVNPFLADLSIDYSSAESVQSSKRLIENLIEPSNKVAITTESSSLTN 354
Oy 342 IERNKHLKTLPMVPEKTKDSPGVSKLGLVLRV-----TRLVNKNKLAIVITRLQN 393
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 355 PEGSEFSPKL-IQSTTKHRGRGMWRQFRLFRKRAMQAFRDGPNNKVAARNSVSA----- 409
Oy 394 LINGELFLFVLRVRSNVLKGAIDRVGLLYQFGAIPYTGMLNANVLPVLRAVSQDS 453
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 410 -----IIFGVFMRMKTQTSIDRMGLQVTAINTAMALITKVGVFPEKERAIVDER 463
Oy 454 QDGLYOKMQLAVLALVLPFSVAVTMIFSSVCYWTGLHPEVARFGVFSALLAPHLIG 513
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 464 AKGSYALGPYLSKLLAEIPIGAAFPILFGSILIPMSKLPFPFSFNAFCGIIVYESFA 523
Oy 514 EFLVLVLGIVQNPINVSVALLSIAGVLVSGFLRNIOEMPPIFKIISYTFQKCYSE 573
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 524 SAMGLYGCAMAPTEAAMALGPSLMTVFIVFG-GYVNPDMTPVIFRMIPRVSILRMAFQ 582
Oy 574 ILVYNEFYGLNF 585
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 583 GLCINEFKGLQF 594
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
Oy 09L182 PRELIMINARY; PRT; 672 AA.
ID 09L182;
AC 09L182;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE ABC transporter-like protein.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA MEDLINE-20363099; PubMed-10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety pl,
RT TAC and BAC clones."
RL DNA Res. 7:217-221(2000).
RL EMBL: AP001313; BAB03081.1;
DR InterPro: IPR003593; AAA_Atpase.
DR InterPro: IPR003439; ABC_transportr.
DR InterPro: IPR003880; Ppantenn_attach.
DR Pfam: PF00005; ABC_tran.1.
DR ProDom: PD000006; ABC_transportr; 1.
DR SMART: SM00382; AAA.1.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN.1.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN.1.
KW ATP-binding.
SQ SEQUENCE 672 AA: 75269 MW: 2082099215600135 CRC64;

```

```

Query Match 18.6%; Score 618.5; DB 10; Length 672;
Best Local Similarity 32.1%; Pred. No. 9.8e-37;
Matches 182; Conservative 100; Mismatches 206; Indels 79; Gaps 17;

Oy 66 ROLLKDVSLVESQINCIISSGSKTTLLDANSGRKRGRTFLGEYVNGR----ALR 121
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 97 RLVIKCVSGIYKPEELAMLALPPSGSKTTLVTALAGRL--QKLSGTIVSYGEPFTSSK 154
Oy 122 RECFQDCEFSYVLOSDTLSSLTVEETLHYTALLAIRGNPSPQK-----VEAVMELS 176
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 155 RKT-----GEVTDQDVLYPHLVTMETLTYTLALRL-----PELTKREKLEVEAVVSDLG 205
Oy 177 LSHVADRLICNYSLGISGTGERRRVINAQLLODPKVALFDEPTTGDCMTANOIVLV 236
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 206 LTRCNSYVIGGLRGISGGERKRVSIQLEKLVNPSILLDEPISGLDSTTAANIVATLR 265
Oy 237 ELARNRNVLTITIQPSSELFQLEDKTAISFGELIFCGTGAEMIDFPNDCGY-PCPEHS 295
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 266 SLARGRTVYTTIHQPSRLRYRMDPKVLYLSEGGPIYSGDSGRVMEYFGSIGYOPGSSFV 325
Oy 296 NPDPFYMDLTSVDOSKER--EIEFSKRVQME-----SAKRSALCHKTKNIE 343
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 326 NPADFVLDLNGITSDTKQYDIETNGRLDLEONSYSKGLSSYKKN--LYPLKE-- 381
Oy 344 RKKHLKTLPMVPEKTKDSPGV-----ESKLGVLRLRVTRNLVRNKLAIVT 388
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 382 --EVSRTFP-----QDQTNARLRKKAITNRPMTSMNQFVLLK---RGLKERSHESFS 430
Oy 389 RLQNLINGELFLFVLRVRSNVLKGAIDRVGLLYQFGAIPYTGMLNANVLPVLRAV 448
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 431 GLRIFMVWSVSLGLLMMHSRV--AHLQDQVGLLFFFSIFMGFFPLNAIFTPQDERPM 488
Oy 449 SDOESQGLYOKMQLAVLALVLPFSVAVTMIFSSVCYWTGLHPEVARFGVFSALLA 508
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 489 LIKERSSTIVRLSSYIARTVGDLPMEILPTITVYTMWGLKPSLTTIMTIVLY 548
Oy 509 PHLICEFLTVLLGIVQNPINVSVALLSIAGVLVSGFLRNIOEMPPIFKIISYTFQ 568
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 549 NVLVAQVGLALGAILMDAKRAATLSVLMVFLLAGY---IQHIDPFIAMLKYSFS 605
Oy 569 KYCEILVYNEFYGLNF-----CGS 589
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 606 HYCYKLV-----GVQYTWDEVYEGGS 627
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
Oy 09FNBS PRELIMINARY; PRT; 727 AA.
ID 09FNBS;
AC 09FNBS;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE ABC transporter-like protein.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA MEDLINE-98069011; PubMed-9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned pl clones."
RL DNA Res. 4:291-300(1997).
RL EMBL: AB006704; BAB08684.1;
DR InterPro: IPR003593; AAA_Atpase.
DR InterPro: IPR003439; ABC_transportr.
DR Pfam: PF00005; ABC_tran.1.
DR ProDom: PD000006; ABC_transportr; 1.

```


Db 508 MARYASV--PNEMLGLITAGLIGIIMTSGFFRLPDLKIFWVPVSYISGSMIAO 565
Qy 571 CSEILVNEFYGLNFTC 587
Db 566 AS-LINGSKAGESRTC 581
RESULT 14
ID 09M2V7 PRELIMINARY; PRT; 720 AA.
AC 09M2V7;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE ABC transporter-like protein.
GN T15C9_80.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucots II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Barques M., Collado M.C., Navarro P., Terol J., Perez-Alonso M.,
RA Mewes H.W., Ruid S., Lemcke K., Mayer K.F.X., Queller F.,
RA Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA eu Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AL132970; CAB82704.1;
DR InterPro: IPR003593; AAA_Arpase.
DR Pfam: PF00005; ABC_tran.1.
DR ProDom: PD000006; ABC_transportr; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KW ATP-binding.
SQ SEQUENCE 720 AA; 79889 MW; 90D5431098CF17A3 CRC64;
Query Match 18.5%; Score 614; DB 10; Length 720;
Best Local Similarity 28.7%; Pred. No. 2,3e-36;
Matches 182; Conservative 122; Mismatches 254; Indels 76; Gaps 17;
Qy 23 SLEEG--APATPEHSLGIIHASYVS-----HRRPMDITSCQOQTRQILKQVS 73
Db 40 SLDSDNDHLMRPVFLSPNNLYTVSVRRKLDPHDLVPRNRRTSFSK---TKTLDNIS 96
Qy 74 LVESGOIMCIIIGSSGKTTLLDMSGRIRAGTFLGCVYVNGFALRQDFQDFSVYL 133
Db 97 GETRGELAVLGAGSGSKTILDLAKRIK--GSLKTVTLNGALOSRMKVISAYVM 155
Qy 134 QSDTLLSLVRETLHYTALAIRNGSGSFQK--KVEAVMAELSHVADRLIGNYSIGG 192
Db 156 QDDLFPMILVEETLMFAEFRLPRSLPKSKKLRLVQALLDQLRNNAKTIIGEGHKG 215
Qy 193 ISTGRRRVSTIAOLDDPKWMLDEPTTGDCMTANOIVLVLELARNRIVVLTTHOP 252
Db 216 ISGGERRVSTIGIDIHPIYVLFDEPTSGIDSTSAFVWVVLKRIASSGSIIMSHIP 275
Qy 253 RSELFQLEDAIISFGLFCGPAEMLDEFNDCGYCPHSPDFDYMLTS----- 306
Db 276 SHRLSLDLRLIFLSRGHTVSSGSPASLPSTFAGNGIPENENOTEPALDLBELGSA 335
Qy 307 -----VDTSKEREIETSKFVOMIESAYKSAICHKTLK-----NIERMHLK----- 349
Db 336 GGTGGLVLEFNKKQEMKQSNPQTLTPASPMP--NLTLMKAISASISRGVLGGGGGS 393
Qy 350 -----TLPWVEFKTKDSGCVSKGLVLRVTRNLVRNKLAVITRLLQNLIMGLFL 401
Db 394 SVINHGQTLAVPAFAMP-----FWIEIKTLTRRSILNSRROPGLIGRLAVIYTG--FIL 448

Qy 402 FVILRVRSNVLKGIADIRYGLXQFVGATPTTCMLNANVLPVLRVAVSDQESODGLYOKM 461
Db 449 ATVEWRIDNSPKG--VOERUG--FFAFAMSTMEFTYCAQALPVLQERITFMHETAYNARRS 506
Qy 462 QMMLAYALVLPFSVATMIFSSVCYWTGCLHEVARFGVSAALLAPHLIGFELTVLL 521
Db 507 SYLSHAIYTFPBLIFLSLAFATTTWANGVLEGGLMGLFYLCLITLIASFSWSSSEVFTLS 566
Qy 522 GIVQNPNIY---NSVALLSINGVLVSGFLRNQIEMPIFKIISYFTFOKYCSEILVYN 578
Db 567 GVV--PHVMIGYIVAIL--AVELLFSGFIRNRDIPQWIMFHLISLKYFEAVLQ 622
Qy 579 EFGVLFNFTGSSNVSTTTPMCAFTGICDIEKT 612
Db 623 EF-----SDTECFVRGVOLFQMS 641
RESULT 15
ID 09ZU35 PRELIMINARY; PRT; 725 AA.
AC 09ZU35;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative ABC transporter.
GN AT2G01320.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN--CV. COLUMBIA;
RC MEDLINE-20083487; PubMed-10617197;
RX Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buehl C.R., Ketchum K.A., Lee J.J., Ronald K.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umeyam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Merz W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RA *Sequence and analysis of chromosome 2 of the plant Arabidopsis
RA thaliana.;
RL Nature 402:761-768 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN--CV. COLUMBIA;
RC Lin X.;
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AC006200; AAD14532.1;
DR InterPro: IPR003593; AAA_Arpase.
DR Pfam: PF00005; ABC_tran.1.
DR ProDom: PD000006; ABC_transportr; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transport.
SQ SEQUENCE 725 AA; 78899 MW; 7DB2E556FE353D7 CRC64;
Query Match 18.4%; Score 610.5; DB 10; Length 725;
Best Local Similarity 29.7%; Pred. No. 4,1e-36;
Matches 166; Conservative 111; Mismatches 244; Indels 37; Gaps 11;
Qy 51 VRP-----WMDTSC-----RQWTRQILKDVLSVYESGOIMCIIIGSSGKTTLLDMSGR 102
Db 65 IRPYTIRNRNITCSLSDSKSKSVFLKKNVSGEAKPGLAIMPSSGKTTLLNVLACQ 124
Qy 103 LGRAGT--FLGEVYVNGFALRQDFQDFSVLOSITLSSVTRVRETLHYTALAIR--G 159
Db 125 LSLSPRLHLSGLLEVNGPSSSKAYK--LAFVROEDLEFFSOLTVRETLSEFAELQLPEIS 182

Wed Jun 11 09:51:14 2003

us-09-989-981a-6.rapb

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 09:12:22 ; Search time 25 Seconds

(without alignments)
2688.377 Million cell updates/sec

Title: US-09-989-981a-6

Perfect score: 3326
Sequence: 1 MGDLSSLPFGSGMGQVNRG.....PALVYLGIVFKIRDLISR 651

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 392085 seqs, 103240269 residues

Total number of hits satisfying chosen parameters: 392085

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database: Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB_PEP:*\n2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB_PEP:*\n3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB_PEP:*\n4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB_PEP:*\n5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB_PEP:*\n6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB_PEP:*\n7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB_PEP:*\n8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB_PEP:*\n9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB_PEP:*\n10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB_PEP:*\n11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB_PEP:*\n12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB_PEP:*\n13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB_PEP:*\n14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB_PEP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3326	100.0	651	US-10-090-455-6	Sequence 6, Appl1
2	3326	100.0	651	US-09-989-981a-6	Sequence 6, Appl1
3	3326	100.0	651	US-09-837-992-3	Sequence 3, Appl1
4	3326	100.0	651	US-09-989-981a-2	Sequence 2, Appl1
5	3326	100.0	651	US-09-837-992-1	Sequence 1, Appl1
6	697	21.0	673	US-09-989-981a-4	Sequence 4, Appl1
7	697	21.0	673	US-10-090-455-7	Sequence 7, Appl1
8	697	21.0	673	US-09-989-981a-8	Sequence 8, Appl1
9	682.5	20.5	655	US-09-981-086-1	Sequence 1, Appl1
10	682.5	20.5	655	US-09-981-353-35	Sequence 35, Appl1
11	680.5	20.5	655	US-10-120-687-61	Sequence 61, Appl1
12	674.5	20.3	655	US-10-090-455-5	Sequence 5, Appl1
13	674.5	20.3	655	US-09-866-866a-10	Sequence 10, Appl1
14	672.5	20.2	655	US-09-866-866a-27	Sequence 27, Appl1
15	660	19.8	657	US-09-866-866a-14	Sequence 14, Appl1
16	602.5	18.1	663	US-10-108-605-245	Sequence 245, App
17	598.5	18.0	674	US-10-090-455-4	Sequence 4, Appl1
18	590.5	17.8	638	US-10-072-621-10	Sequence 10, Appl1
19	585.5	17.6	646	US-10-072-621-9	Sequence 9, Appl1

20	585.5	17.6	646	US-10-090-455-2	Sequence 2, Appl1
21	578.5	17.4	637	US-10-090-455-8	Sequence 8, Appl1
22	578	17.4	604	US-09-745-763-197	Sequence 197, App
23	570.5	17.2	646	US-10-154-452-4	Sequence 4, Appl1
24	567.5	17.1	646	US-10-090-455-13	Sequence 13, Appl1
25	562.5	16.9	646	US-10-154-452-8	Sequence 8, Appl1
26	485	14.6	545	US-10-083-357-1335	Sequence 1335, App
27	403.5	12.1	1564	US-09-801-368-244	Sequence 244, App
28	398	12.0	1501	US-09-801-368-346	Sequence 346, App
29	374	11.2	1511	US-09-801-368-250	Sequence 250, App
30	270.5	8.1	306	US-09-738-626-4554	Sequence 4554, App
31	268	8.1	242	US-09-769-787-83	Sequence 83, Appl1
32	263.5	7.9	345	US-09-815-242-11068	Sequence 11068, A
33	263	7.9	241	US-10-097-340-2	Sequence 1194, A
34	241.5	7.3	247	US-10-097-340-2	Sequence 5108, Ap
35	240	7.2	1279	US-10-097-340-2	Sequence 2, Appl1
36	240	7.2	1280	US-09-866-866a-2	Sequence 7, Appl1
37	240	7.2	1280	US-09-866-866a-4	Sequence 4, Appl1
38	240	7.2	1280	US-10-156-239-8	Sequence 8, Appl1
39	239.5	7.2	2436	US-10-199-485-8	Sequence 8, Appl1
40	239.5	7.2	2436	US-09-795-693-8	Sequence 8, Appl1
41	239.5	7.2	2436	US-09-971-121-4	Sequence 4, Appl1
42	238.5	7.2	1594	US-09-971-121-2	Sequence 2, Appl1
43	238.5	7.1	1642	US-09-769-787-144	Sequence 144, App
44	236.5	7.1	385	US-10-252-819-17	Sequence 17, Appl1
45	236	7.1	203	US-10-252-819-17	

ALIGNMENTS

RESULT 1				
US-10-090-455-6				
Sequence 6, Application US/10090455				
Publication NO. US20030027259A1				
GENERAL INFORMATION				
APPLICANT: Chen, Hongyun				
APPLICANT: Le Bihan, Stephane				
TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF				
FILE REFERENCE: 100103.406				
CURRENT APPLICATION NUMBER: US/10/090,455				
CURRENT FILING DATE: 2002-03-01				
NUMBER OF SEQ ID NOS: 17				
SOFTWARE: FASTSEQ for Windows Version 4.0				
SEQ ID NO 6				
LENGTH: 651				
TYPE: PRT				
ORGANISM: Homo sapiens				
US-10-090-455-6				
Query Match				
Best Local Similarity 100.0%; Score 3326; DB 9; Length 651;				
Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
OY	1	MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSLGILHASTVSHRRPMDITSC	60	
DB	1	MGDLSSLPFGSGMGQVNRGSSLEGAPATAPPEPHSLGILHASTVSHRRPMDITSC	60	
OY	61	ROOMTRQILKDVSLVSGOIMCITGSSGKTTLLDMSGRLGAGTFLGYYNGRL	120	
DB	61	ROOMTRQILKDVSLVSGOIMCITGSSGKTTLLDMSGRLGAGTFLGYYNGRL	120	
OY	121	RREDFQDFSVYVLOSSTLSTLVTRETLHYTALAIRGNPGSPFKTEAVAAETSLSHV	180	
DB	121	RREDFQDFSVYVLOSSTLSTLVTRETLHYTALAIRGNPGSPFKTEAVAAETSLSHV	180	
OY	181	ADRLIGNTSLGISTGERRRYSIAAQLQDPKVMLEDEPTTGLDCTMTANOIVLLVELAR	240	
DB	181	ADRLIGNTSLGISTGERRRYSIAAQLQDPKVMLEDEPTTGLDCTMTANOIVLLVELAR	240	
OY	241	RNRIVYLTTHOPRSLFQPKIAITLNGELIFGTPAEMDDEFNDCCYPCPEHNPDPF	300	
DB	241	RNRIVYLTTHOPRSLFQPKIAITLNGELIFGTPAEMDDEFNDCCYPCPEHNPDPF	300	
OY	241	RNRIVYLTTHOPRSLFQPKIAITLNGELIFGTPAEMDDEFNDCCYPCPEHNPDPF	300	
DB	241	RNRIVYLTTHOPRSLFQPKIAITLNGELIFGTPAEMDDEFNDCCYPCPEHNPDPF	300	

OY 301 YMDLSDVDPQSKEREIETSKRVQMIESAKKSAICHKTLKNIERKHLKTLPMVPEKTKD 360
DB 301 YMDLSDVDPQSKEREIETSKRVQMIESAKKSAICHKTLKNIERKHLKTLPMVPEKTKD 360
OY 361 SPGVSKLGVLLRRTRRLVRRKLAIVITRLONLIMGFLFELFVLRVSNVTKGAIODRV 420
DB 361 SPGVSKLGVLLRRTRRLVRRKLAIVITRLONLIMGFLFELFVLRVSNVTKGAIODRV 420
OY 421 GLLYOFVATPYTGMLNANVLEPVLRAVSDOSDGLVOKQOMMLATVLAHVLPESVATM 480
DB 421 GLLYOFVATPYTGMLNANVLEPVLRAVSDOSDGLVOKQOMMLATVLAHVLPESVATM 480
OY 481 IFSSVCYWTGLHPEVARGYFSALLAPHLIGFELTVLGLGVONPNVNSVALLSIA 540
DB 481 IFSSVCYWTGLHPEVARGYFSALLAPHLIGFELTVLGLGVONPNVNSVALLSIA 540
OY 541 GVLVSSGFLRNIOEMPFPKIIISYTFPOKYCEIIVNEFGLNFTCGSSNVSVTTPMC 600
DB 541 GVLVSSGFLRNIOEMPFPKIIISYTFPOKYCEIIVNEFGLNFTCGSSNVSVTTPMC 600
OY 601 AFTGIOFTEKTCPGATSRFTMNFILISFIPALVILGIVVKIRIDHLISR 651
DB 601 AFTGIOFTEKTCPGATSRFTMNFILISFIPALVILGIVVKIRIDHLISR 651

RESULT 2

US-09-989-981A-6
Sequence 6, Application US/09989981A
Publication No. US20030049730A1
GENERAL INFORMATION:

APPLICANT: Hobbs, Helen H.
APPLICANT: Shan, Bel
APPLICANT: Barnes, Robert
APPLICANT: Tian, Hui
APPLICANT: Tularik Inc.
TITLE OF INVENTION: Board of Regents, The University of Texas System
FILE REFERENCE: 018781-007320US
CURRENT APPLICATION NUMBER: US/09/989,981A
CURRENT FILING DATE: 2002-07-23
PRIOR APPLICATION NUMBER: US 60/252,235
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/253,645
PRIOR FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 13
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 6
LENGTH: 651
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: human ABCG5 (hABCG5)
US-09-989-981A-6

Query Match 100.0%; Score 3326; DB 9; Length 651;
Best Local Similarity 100.0%; Pred. No. 1.6e-289;
Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MGDLSLTPGGSMGLQVNRGSSSLEGAPATAPPEHSGILHASTVSHRRPMMWDTSC 60
DB 1 MGDLSLTPGGSMGLQVNRGSSSLEGAPATAPPEHSGILHASTVSHRRPMMWDTSC 60
OY 61 RQMTROLIKDVSLYVESGOIMCITLSSGSGKTTLLDMSGRIRAGFTLGEVYVNGAL 120
DB 61 RQMTROLIKDVSLYVESGOIMCITLSSGSGKTTLLDMSGRIRAGFTLGEVYVNGAL 120
OY 121 RRFODCFSTYVLOSDTLSSLYRETLHYTALLAIRGNPGSFQKKEAVMAELSLSHV 180
DB 121 RRFODCFSTYVLOSDTLSSLYRETLHYTALLAIRGNPGSFQKKEAVMAELSLSHV 180
OY 181 ADRLLGNSTLGGISGERRRYSIAQLLODPKVMFLDEPTGLDCMTANQIVLLVELAR 240
DB 181 ADRLLGNSTLGGISGERRRYSIAQLLODPKVMFLDEPTGLDCMTANQIVLLVELAR 240

DB 181 ADRLLGNSTLGGISGERRRYSIAQLLODPKVMFLDEPTGLDCMTANQIVLLVELAR 240
OY 241 RRRIVVLTIHOPRSELFOLFDRKIALISFGEILFCTPAMELDFPNDCGCPCEHSPNPF 300
DB 241 RRRIVVLTIHOPRSELFOLFDRKIALISFGEILFCTPAMELDFPNDCGCPCEHSPNPF 300
OY 301 YMDLSDVDPQSKEREIETSKRVQMIESAKKSAICHKTLKNIERKHLKTLPMVPEKTKD 360
DB 301 YMDLSDVDPQSKEREIETSKRVQMIESAKKSAICHKTLKNIERKHLKTLPMVPEKTKD 360
OY 361 SPGVSKLGVLLRRTRRLVRRKLAIVITRLONLIMGFLFELFVLRVSNVTKGAIODRV 420
DB 361 SPGVSKLGVLLRRTRRLVRRKLAIVITRLONLIMGFLFELFVLRVSNVTKGAIODRV 420
OY 421 GLLYOFVATPYTGMLNANVLEPVLRAVSDOSDGLVOKQOMMLATVLAHVLPESVATM 480
DB 421 GLLYOFVATPYTGMLNANVLEPVLRAVSDOSDGLVOKQOMMLATVLAHVLPESVATM 480
OY 481 IFSSVCYWTGLHPEVARGYFSALLAPHLIGFELTVLGLGVONPNVNSVALLSIA 540
DB 481 IFSSVCYWTGLHPEVARGYFSALLAPHLIGFELTVLGLGVONPNVNSVALLSIA 540
OY 541 GVLVSSGFLRNIOEMPFPKIIISYTFPOKYCEIIVNEFGLNFTCGSSNVSVTTPMC 600
DB 541 GVLVSSGFLRNIOEMPFPKIIISYTFPOKYCEIIVNEFGLNFTCGSSNVSVTTPMC 600
OY 601 AFTGIOFTEKTCPGATSRFTMNFILISFIPALVILGIVVKIRIDHLISR 651
DB 601 AFTGIOFTEKTCPGATSRFTMNFILISFIPALVILGIVVKIRIDHLISR 651

RESULT 3

US-09-837-992-3
Sequence 3, Application US/09837992
Patent No. US20020081687A1
GENERAL INFORMATION:

APPLICANT: Tian, Hui
APPLICANT: Schultz, Joshua
APPLICANT: Shan, Bel
APPLICANT: Tularik Inc.
TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions
FILE REFERENCE: 018781-006020US
CURRENT APPLICATION NUMBER: US/09/837,992
CURRENT FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: US 60/198,465
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: US 60/204,234
PRIOR FILING DATE: 2000-05-15
NUMBER OF SEQ ID NOS: 45
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 3
LENGTH: 651
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: human sitosterolemia susceptibility gene (SSG)
US-09-837-992-3

Query Match 100.0%; Score 3326; DB 10; Length 651;
Best Local Similarity 100.0%; Pred. No. 1.6e-289;
Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MGDLSLTPGGSMGLQVNRGSSSLEGAPATAPPEHSGILHASTVSHRRPMMWDTSC 60
DB 1 MGDLSLTPGGSMGLQVNRGSSSLEGAPATAPPEHSGILHASTVSHRRPMMWDTSC 60
OY 61 RQMTROLIKDVSLYVESGOIMCITLSSGSGKTTLLDMSGRIRAGFTLGEVYVNGAL 120
DB 61 RQMTROLIKDVSLYVESGOIMCITLSSGSGKTTLLDMSGRIRAGFTLGEVYVNGAL 120
OY 121 RRFODCFSTYVLOSDTLSSLYRETLHYTALLAIRGNPGSFQKKEAVMAELSLSHV 180
DB 121 RRFODCFSTYVLOSDTLSSLYRETLHYTALLAIRGNPGSFQKKEAVMAELSLSHV 180

```

Db      121 REEQPQDFSVYLOSDFLLSLTIVETLHTLALIRKNGSSFOKKEAVAAELSLNV 180
Oy      181 ADRLIGNTSLGISTGERRRVSIAAQLLODPKVMLEFDEPTTGDCMTANOAVLVLELAR 240
Db      181 ADRLIGNTSLGISTGERRRVSIAAQLLODPKVMLEFDEPTTGDCMTANOAVLVLELAR 240
Oy      241 RNRIVYLTIHOPRSELFOPLDKIALISFGELICGTPAEMLDFNDCGYPCEHSNPDF 300
Db      241 RNRIVYLTIHOPRSELFOPLDKIALISFGELICGTPAEMLDFNDCGYPCEHSNPDF 300
Oy      301 YMDLTSVDQSKEREIETSKRVOMIESAYKKSACHKTKNIERKHLKTLPMVPFKTKD 360
Db      301 YMDLTSVDQSKEREIETSKRVOMIESAYKKSACHKTKNIERKHLKTLPMVPFKTKD 360
Oy      361 SPGVFSKLGVLRRVTRNLVANKLAVITRLQNLIMGLFELFVLRRVSNVLKGAIDQDR 420
Db      361 SPGVFSKLGVLRRVTRNLVANKLAVITRLQNLIMGLFELFVLRRVSNVLKGAIDQDR 420
Oy      421 GLLYOFVGAFTPTGMLNAVNFVPLRAVSQDESQDGLYQKQMMALAVLHVPFSVATM 480
Db      421 GLLYOFVGAFTPTGMLNAVNFVPLRAVSQDESQDGLYQKQMMALAVLHVPFSVATM 480
Oy      481 IFSSVCYWTGLHPEVARFGEFSALLAPHLIGFELTVLVLGIYQNPRTVNSVVALLSIA 540
Db      481 IFSSVCYWTGLHPEVARFGEFSALLAPHLIGFELTVLVLGIYQNPRTVNSVVALLSIA 540
Oy      541 GVLVSGFLRNIOEMPIPKIISYFTQKCYSEILVYNEFGNFTGSSNVSATTPMNC 600
Db      541 GVLVSGFLRNIOEMPIPKIISYFTQKCYSEILVYNEFGNFTGSSNVSATTPMNC 600
Oy      601 AFQOGIOFIETKTCGATSRFTANFLIYFTIPALVILGIYVEKIRDLISR 651
Db      601 AFQOGIOFIETKTCGATSRFTANFLIYFTIPALVILGIYVEKIRDLISR 651

```

RESULT 4

US-09-989-981A-2

Sequence 2, Application US/09989981A

Publication No. US20030049730A1

GENERAL INFORMATION:

```

APPLICANT: Hobbs, Helen H.
APPLICANT: Shan, Bel
APPLICANT: Barnes, Robert
APPLICANT: Tian, Hui
APPLICANT: Tularik Inc.
TITLE OF INVENTION: Board of Regents, The University of Texas System
FILE REFERENCE: 018781-00732005
CURRENT APPLICATION NUMBER: US/09/989,981A
CURRENT FILING DATE: 2002-07-23
PRIOR APPLICATION NUMBER: US 60/252,235
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/253,645
PRIOR FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 652
TYPE: PRT
ORGANISM: Mus musculus
FEATURE:
OTHER INFORMATION: mouse ABCG5 (mABCG5)
US-09-989-981A-2

```

```

Query Match      82.5%; Score 2744.5; DB 9; Length 652;
Best Local Similarity 80.2%; Pred. No. 2.2e-237;
Matches 523; Conservative 64; Mismatches 64; Indels 1; Gaps 1;
Oy      1 MGDLSLTPGGSMGIQVNGRSGSSLEGAPATAPDP-HSIIHIAASYSVSHRRPRMDITS 59
Db      1 MGEPLFLSPGARGHINRSGSSLEGAPATAPDP-HSIIHIAASYSVSHRRPRMDITS 60

```

```

Oy      60 CROQWTRQILKDVSLYVESGOIMCITGSSGCKTLLDAMSGRLGACTFLGEVYNGRA 119
Db      61 COQKMDROLKDVSLYIESGOIMCITGSSGCKTLLDAMSGRLGACTFLGEVYNGRA 120
Oy      120 LREQFQDFSVYLOSDFLLSLVRETLYHTALAIRKNGSSFOKKEAVAAELSLNV 179
Db      121 LREQFQDFSVYLOSDFLLSLVRETLYHTALAIRKNGSSFOKKEAVAAELSLNV 180
Oy      180 VADRLIGNTSLGISTGERRRVSIAAQLLODPKVMLEFDEPTTGDCMTANOAVLVLELAR 239
Db      181 VADRLIGNTSLGISTGERRRVSIAAQLLODPKVMLEFDEPTTGDCMTANOAVLVLELAR 240
Oy      240 RNRIVYLTIHOPRSELFOPLDKIALISFGELICGTPAEMLDFNDCGYPCEHSNPDF 299
Db      241 RNRIVYLTIHOPRSELFOPLDKIALISFGELICGTPAEMLDFNDCGYPCEHSNPDF 300
Oy      300 FYMDLTSVDQSKEREIETSKRVOMIESAYKKSACHKTKNIERKHLKTLPMVPFKTK 359
Db      301 FYMDLTSVDQSKEREIETSKRVOMIESAYKKSACHKTKNIERKHLKTLPMVPFKTK 360
Oy      360 DSPGVFSKLGVLRRVTRNLVANKLAVITRLQNLIMGLFELFVLRRVSNVLKGAIDQDR 419
Db      361 DSPGVFSKLGVLRRVTRNLVANKLAVITRLQNLIMGLFELFVLRRVSNVLKGAIDQDR 420
Oy      420 VGLLYOFVGAFTPTGMLNAVNFVPLRAVSQDESQDGLYQKQMMALAVLHVPFSVAT 479
Db      421 VGLLYOFVGAFTPTGMLNAVNFVPLRAVSQDESQDGLYQKQMMALAVLHVPFSVAT 480
Oy      480 MIFSSVCYWTGLHPEVARFGEFSALLAPHLIGFELTVLVLGIYQNPRTVNSVVALLSI 539
Db      481 MIFSSVCYWTGLHPEVARFGEFSALLAPHLIGFELTVLVLGIYQNPRTVNSVVALLSI 540
Oy      540 AGVLVSGFLRNIOEMPIPKIISYFTQKCYSEILVYNEFGNFTGSSNVSATTPMNC 599
Db      541 AGVLVSGFLRNIOEMPIPKIISYFTQKCYSEILVYNEFGNFTGSSNVSATTPMNC 600
Oy      600 CAFQOGIOFIETKTCGATSRFTANFLIYFTIPALVILGIYVEKIRDLISR 651
Db      601 CAFQOGIOFIETKTCGATSRFTANFLIYFTIPALVILGIYVEKIRDLISR 652

```

RESULT 5

US-09-837-992-1

Sequence 1, Application US/09837992

Patent No. US20020081687A1

GENERAL INFORMATION:

```

APPLICANT: Tian, Hui
APPLICANT: Schultz, Joshua
APPLICANT: Shan, Bel
APPLICANT: Tularik Inc.
TITLE OF INVENTION: SLC12A1, a gene involved in the regulation of intracellular sodium concentration
FILE REFERENCE: 018781-00602005
CURRENT APPLICATION NUMBER: US/09/837,992
CURRENT FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: US 60/198,465
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: US 60/204,234
PRIOR FILING DATE: 2000-05-15
NUMBER OF SEQ ID NOS: 45
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 652
TYPE: PRT
ORGANISM: Mus musculus
FEATURE:
OTHER INFORMATION: mouse SLC12A1, a gene involved in the regulation of intracellular sodium concentration
US-09-837-992-1

```

```

Query Match      82.5%; Score 2744.5; DB 10; Length 652;
Best Local Similarity 80.2%; Pred. No. 2.2e-237;
Matches 523; Conservative 64; Mismatches 64; Indels 1; Gaps 1;
Oy      1 MGDLSLTPGGSMGIQVNGRSGSSLEGAPATAPDP-HSIIHIAASYSVSHRRPRMDITS 59
Db      1 MGEPLFLSPGARGHINRSGSSLEGAPATAPDP-HSIIHIAASYSVSHRRPRMDITS 60

```

Query Match	21.0%;	Score 697;	DB 9;	Length 672;
Best Local Similarity	29.1%;	Pred. No. 9.4e-54;		
Matches	195;	Conservative 129;	Mismatches 265;	Indels 84; Gaps 18
15	LQVNRGSSSSLEGAPATAPPEPHSLGILHASTVSVHRV-----	PWMD-ITSCR	61	
17	LQDASGLDSDL-----	FSSESNDNSTLYFTVSGSNTLEVVDLTYQVYDIASQVPMFEQDAQPK	72	
62	QOQWTRQI-----	LKQVSLYVEGQIMCIGLSSGSKRTILLDAMSGRLGRAGTF-LGE	112	
73	IPMRSHSSODSCGELGIRNLSFRKVRNGQIMATLISGSGCRASLDYITGR-GHGQKKSQ	131		
113	VYVNGRALRRBQFODCFSTYVSLQSPDLSLTYRETLHTALLAI-RRGNPSPFOKVEAY	171		
132	IWINQPTPTQYLRKCAVHARQHOQLPNTLYRETLAFIAQRLPRTFSQAQRDRKVEDV	191		
172	MAELSLSHVADRLIGNTSLSGISTGERRRVSIQAQLQDPRVMLFDEPTTGIDCCTANQI	231		
192	IAELRLQCAMTRVNGTIVYRGVSGGERRRVSIQVQLNPGILLIDETSGIDSTPTANL	251		
232	VVLVELLARRRRIYVLTIHQPRSELPOLFKAIALISFSELIFCGTPRAMLDFPNDGTPC	291		
252	VTTLSRLAKGRNLVLSLHQPRSDIFRFLDLVLLMTSGTPYILGMAQOMVOYFTSIGHPC	311		
292	PEHSPDFEYDLSVDTSQKEREIETSKRYOMIESAKKSA-----	ICRKLTKNIERM	345	
312	PRTPNPADFEYDLSIDRRSKEREYATYERKAQSLALFLLEVQGFDFELMKRAEAKELMTS	371		
346	KHLTLPMWPKRTDS-----	GPVPSKLGVLILRYTRMLVRRKLAVITRRLQNLQNG	397	
372	THVYSITL-----	TQDTCGTAVELPQMLEQSTILRRQISNDFRDLPTLLIHGSEACLAS	427	
398	LFLLEFVLRRVSNVLKGAIDRVGLIQVFGATPYTGMLNVLNPEVYLRAVSDQESDGL	457		
428	LILGFLYXGHGAKQL-SFMDPTALFLMIGALIPFNVLIDVVSCHSRSMLYETLEGL	485		
458	YQKQMMALVLAHLTPFSVVAITMFSSCYTTLGLHPEVAFGFSALLAPHLIGELV-	516		
486	YTAGPYFPFAKLIGELPEHCAYVLIITAMPYVLTMLRPPELF-----	LT--HFLVWLIV	537	
517	-----	TVLLIGIVQNPNT-VNSVALLSIAGVLVGSGLVIOEMPIPKLISYFTPOK	570	
538	VECCRTALALASAMLPFFHMSFPCNALYNSFTYTAGFMALMDMLMIVANISKLSFLRW	597		
571	CSEILVNEFEYGLNFT--CGSSNVSVTTPMCAFTQIOGLEKTCPGATSRFTNFFLXY	628		
598	CFSGLMOIQFNGHLTYTQIGNFTFSILGDTM-----	ISANDNSHLY	640	
629	SFIPALVITLGI	639		
641	ATV--LIVIGI	649		

Wed Jun 11 09:51:14 2003

us-09-989-981a-6.rapb

Page 5

Query Match 21.0%; Score 697; DB 9; Length 673;
Best Local Similarity 28.9%; Pred. No. 9, 5e-54;
Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;

```
OY 8 TPGSGMGLOVNRGSSSLEGPAT--APEPHSIGILHASYSHRR--PMD--ITSROOM 64
DB 16 TPQDTSGLODRLFSSSDNSLYFTYSGQNTLEVDNLOVLAQVPEFEOIAQFKMFW 75
OY 65 TROI-----LKQVSLVESGQIMCIIIGSSGSKTTLIDAMSGRLAGTF--LGEVYV 115
DB 76 TSPSCNSCELCIOMISFVRSGOMALIIIGSSGCRASHLDVITGR--GHGKIKSGQIWI 134
OY 116 NGRALRRQFODCFEYVLOSOTLLSLTYRETLHTALLAT--RGNPGSFORKVEAVMAE 174
DB 135 NGQSSPOLVRKCAVHROHNOLEPNLYRETLIAQWRLEPRTSOAQRKREVDIAE 194
OY 175 LSLSHVADRLIGNYSIGISTGERRRVSIAQLODPKMLEDEPTTGIDCMTANOIYVL 234
DB 195 LRLRCADTRVGNMVRGLSGERRRVSIGVOLLNPGILILDEPTSGIDSTAHNLVKT 254
OY 235 LVELARRRRIYVLTTHQPRSELFDKIAISFGELIFCGTPAEMLDFPNDGYPCEH 294
DB 255 LSLRAGNRVLVLSIHQPRSDIFRFLDLVLTMTSGTPIYLGAAQMVQYFAIGYPCRY 314
OY 295 SNPDFYMDLTSVDSQKEREIETSKRVOMIESAKKSAICHKTLKNIERMKHL----- 348
DB 315 SNPADFYDLTSIDRRSREDELAEREKAQSLAALF-----LEKRDLDDEFLMK 362
OY 349 -----KTLPM-----VPEKTKDSGYSKIGVLLRVRTRLVNRKLAIVTRL 390
DB 363 AETKDLDDETCVSSVPTLDNCLPSPTK--MGAQOQTTILIRDISDFDLPTLLHG 421
OY 391 LQNLINGLELFFVLRRNSNVLKGAIQ---DRGLLYQVFGATPYTGMLNANVLEPYLR 446
DB 422 AEACIMSWTIGFLYFG-----HGSIOISFMDTALLPMICALIPFNVLIDVISCYSER 475
OY 447 AVSQOESODGLYQKQOMLALAVLPESVATMIFSSVCYWTGLHEVARF----- 499
DB 476 AMLYTELEDGLYTGPFYFAKILGELPEHCAYIIYGMPTWMLNLRGLQPFLLHFLV 535
OY 500 -----GYFAALLAPHLIGEFLTLVLGIVQONIVNSVALLSTINGVYVSGEL 549
DB 536 WLVEVCCRIMALAALALPFFHMASFPS-----NALYNSYTLAG-----GFM 577
OY 550 RNIDEMPIPEKIIISYTFQKYSSEILVNEFYGLNFTGSSNSVSTN 597
DB 578 INSLMTVPAMISKVSFLKWCFFGLMKIQPSRRTYKMPGLNLTIVAS 625
```

RESULT 8
US-09-989-981a-8
Sequence 8, Application US/09989981A
Publication No. US20030049730A1
GENERAL INFORMATION:
APPLICANT: Hobbs, Helen H.
APPLICANT: Shan, Bei
APPLICANT: Barnes, Robert
APPLICANT: Tian, Hui
APPLICANT: Tularik Inc.
TITLE OF INVENTION: Board of Regents, The University of Texas System
FILE REFERENCE: ABC5 and ABC8: Compositions and Methods of Use
CURRENT APPLICATION NUMBER: US/09/989, 981A
PRIOR FILING DATE: 2002-07-23
PRIOR APPLICATION NUMBER: US 60/252, 235
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/253, 645
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 8
LENGTH: 673

TYPE: PR
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: human ABC8 (hABC8)
US-09-989-981a-8

Query Match 21.0%; Score 697; DB 9; Length 673;
Best Local Similarity 28.9%; Pred. No. 9, 5e-54;
Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;

```
OY 8 TPGSGMGLOVNRGSSSLEGPAT--APEPHSIGILHASYSHRR--PMD--ITSROOM 64
DB 16 TPQDTSGLODRLFSSSDNSLYFTYSGQNTLEVDNLOVLAQVPEFEOIAQFKMFW 75
OY 65 TROI-----LKQVSLVESGQIMCIIIGSSGSKTTLIDAMSGRLAGTF--LGEVYV 115
DB 76 TSPSCNSCELCIOMISFVRSGOMALIIIGSSGCRASHLDVITGR--GHGKIKSGQIWI 134
OY 116 NGRALRRQFODCFEYVLOSOTLLSLTYRETLHTALLAT--RGNPGSFORKVEAVMAE 174
DB 135 NGQSSPOLVRKCAVHROHNOLEPNLYRETLIAQWRLEPRTSOAQRKREVDIAE 194
OY 175 LSLSHVADRLIGNYSIGISTGERRRVSIAQLODPKMLEDEPTTGIDCMTANOIYVL 234
DB 195 LRLRCADTRVGNMVRGLSGERRRVSIGVOLLNPGILILDEPTSGIDSTAHNLVKT 254
OY 235 LVELARRRRIYVLTTHQPRSELFDKIAISFGELIFCGTPAEMLDFPNDGYPCEH 294
DB 255 LSLRAGNRVLVLSIHQPRSDIFRFLDLVLTMTSGTPIYLGAAQMVQYFAIGYPCRY 314
OY 295 SNPDFYMDLTSVDSQKEREIETSKRVOMIESAKKSAICHKTLKNIERMKHL----- 348
DB 315 SNPADFYDLTSIDRRSREDELAEREKAQSLAALF-----LEKRDLDDEFLMK 362
OY 349 -----KTLPM-----VPEKTKDSGYSKIGVLLRVRTRLVNRKLAIVTRL 390
DB 363 AETKDLDDETCVSSVPTLDNCLPSPTK--MGAQOQTTILIRDISDFDLPTLLHG 421
OY 391 LQNLINGLELFFVLRRNSNVLKGAIQ---DRGLLYQVFGATPYTGMLNANVLEPYLR 446
DB 422 AEACIMSWTIGFLYFG-----HGSIOISFMDTALLPMICALIPFNVLIDVISCYSER 475
OY 447 AVSQOESODGLYQKQOMLALAVLPESVATMIFSSVCYWTGLHEVARF----- 499
DB 476 AMLYTELEDGLYTGPFYFAKILGELPEHCAYIIYGMPTWMLNLRGLQPFLLHFLV 535
OY 500 -----GYFAALLAPHLIGEFLTLVLGIVQONIVNSVALLSTINGVYVSGEL 549
DB 536 WLVEVCCRIMALAALALPFFHMASFPS-----NALYNSYTLAG-----GFM 577
OY 550 RNIDEMPIPEKIIISYTFQKYSSEILVNEFYGLNFTGSSNSVSTN 597
DB 578 INSLMTVPAMISKVSFLKWCFFGLMKIQPSRRTYKMPGLNLTIVAS 625
```

RESULT 9
US-09-961-086-1
Sequence 1, Application US/09961086
Publication No. US20030036645A1
GENERAL INFORMATION:
APPLICANT: UNIVERSITY OF MARYLAND, BALTIMORE
APPLICANT: ROSS, Douglas D.
APPLICANT: DOTLE, L. Austin
TITLE OF INVENTION: BREAST CANCER RESISTANCE PROTEIN (BCRP) AND THE DNA
FILE REFERENCE: AB020, Lynne
CURRENT APPLICATION NUMBER: US/09/961, 086
PRIOR FILING DATE: 2001-09-21
PRIOR APPLICATION NUMBER: US 60/073, 763
PRIOR FILING DATE: 1998-02-05
PRIOR APPLICATION NUMBER: PCT/US99/02577
PRIOR FILING DATE: 1999-02-05

NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 655
TYPE: PRT
ORGANISM: Homo sapiens
US-09-961-086-1

Query Match 20.5%; Score 682.5; DB 9; Length 655;
Best Local Similarity 29.2%; Pred. No. 1.8e-52;
Matches 182; Conservative 138; Mismatches 249; Indels 55; Gaps 18;

21 SSSLEGAPATAP---EPHSLGLTHASYSVSHRPRMMDITSCROQMTROILKDVSLYE 77
13 SGNMTGFPATASNDKAFTEGAVLSFHNICYVKLKSGFLPCRKEVEKILLSNINGIMK 72
78 SQGIMCIISSSGSKTLLDAMSGRLGRAGTFLGEYTVNGRARRQPDCCSYVQSOT 137
73 PG-LNALIPPTGGKSSLDVLAARDPSG-LSGDVLLING-APRPANFCNSGYVQDDV 129
138 LLSLTVRETLHTALLAIRGNPG-SFOKVEAVAEISLHVADRLIGNYSIGISGTG 196
130 VNGTLVRENLOPESALRLATMTNHEKNERIRVIOELGDKVADSKYGTOTIRGVSG 189
197 ERRRVSIQAOLLODPKVMLEFDEPTGLDCMTANOIYVLLVELARRNRIVYLTIHQPSRL 256
190 EKRRTSIGMELITDPSILFDEPTGLDSSSTANAVLLLRMSKOGRTIIFSIHQPRYSI 249
257 FOLFXIALISGCELIFCSTPAEMDFNDCCGPRCEHNPDPFYDLTSVDTQ---SK 312
250 FKLPSLTLASGRLEHFGPAQALGYFESAGHCAVANNPADFDLIINDGSTAVALLNR 309
313 ERE-----IETSKR-----VOMIESAYKRSATCHKT-----LKNTERMKHLTPMVPF 356
310 EEDFRATEIIEPSKODKPLIEKLAELIYVSSFYKELKALHOLSGEKKKKTIVREISY 369
357 KTKDSPGVSKGLVLRRTNRLVNRKLAIVITRLLONLIMGFL--LFFVLVRSVNLVG 414
370 TT-----SFCHQLRWMSKRSFKMLGNPQASIAQIITYVVLGVIGAIYGLKNDST---- 421
415 ATODRGILYFVGAPTYGMLNAVLFVYLAVSQESQDGLYQKQOMMLAYAL-HVLP 473
422 GIONRAGVLE-FLTNOCSSVSVALEFVEKRLIHEITISGYRVSFYGLKLSLDLP 480
474 FSVVATMIFSSVCYWTGLGHPVARGFSALLAPHLIGFEFLTVLGLIYONPNIVNSY 533
481 MMLPSIIFTCIYFIMLGKPKADAFVMMFTLM--WVAYSASSMALAIAGOSVSVSA 537
534 VALLSIAGY--LVGSGFLNIQIEMPIPEKTIISYTFQKCYSEILVYNEFYGLNFTCGSSN 591
538 TLMATICVFYFMIFSGLLVNLTTIASWLSMLOFYFISIPRYGFTALQHNFEIAGNFCPG--- 594
592 VSVTNPMCAFTQGIQIEKTCPG 615
595 LNATGNPCNTA-----TCTG 610

RESULT 10
US-09-981-353-35
Sequence 35, Application US/09981353
Patent No. US20020160382A1
GENERAL INFORMATION:
APPLICANT: Lasek, Amy W.
TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
FILE REFERENCE: PA-0038 US
CURRENT APPLICATION NUMBER: US/09/981,353
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PERL Program
SEQ ID NO 35
LENGTH: 655
TYPE: PRT

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20020160382A1 5517972CD1
US-09-981-353-35

Query Match 20.5%; Score 680.5; DB 9; Length 655;
Best Local Similarity 29.2%; Pred. No. 2.7e-52;
Matches 182; Conservative 137; Mismatches 250; Indels 55; Gaps 18;

21 SSSLEGAPATAP---EPHSLGLTHASYSVSHRPRMMDITSCROQMTROILKDVSLYE 77
13 SGNMTGFPATASNDKAFTEGAVLSFHNICYVKLKSGFLPCRKEVEKILLSNINGIMK 72
78 SQGIMCIISSSGSKTLLDAMSGRLGRAGTFLGEYTVNGRARRQPDCCSYVQSOT 137
73 PG-LNALIPPTGGKSSLDVLAARDPSG-LSGDVLLING-APRPANFCNSGYVQDDV 129
138 LLSLTVRETLHTALLAIRGNPG-SFOKVEAVAEISLHVADRLIGNYSIGISGTG 196
130 VNGTLVRENLOPESALRLATMTNHEKNERIRVIOELGDKVADSKYGTOTIRGVSG 189
197 ERRRVSIQAOLLODPKVMLEFDEPTGLDCMTANOIYVLLVELARRNRIVYLTIHQPSRL 256
190 EKRRTSIGMELITDPSILFDEPTGLDSSSTANAVLLLRMSKOGRTIIFSIHQPRYSI 249
257 FOLFXIALISGCELIFCSTPAEMDFNDCCGPRCEHNPDPFYDLTSVDTQ---SK 312
250 FKLPSLTLASGRLEHFGPAQALGYFESAGHCAVANNPADFDLIINDGSTAVALLNR 309
313 ERE-----IETSKR-----VOMIESAYKRSATCHKT-----LKNTERMKHLTPMVPF 356
310 EEDFRATEIIEPSKODKPLIEKLAELIYVSSFYKELKALHOLSGEKKKKTIVREISY 369
357 KTKDSPGVSKGLVLRRTNRLVNRKLAIVITRLLONLIMGFL--LFFVLVRSVNLVG 414
370 TT-----SFCHQLRWMSKRSFKMLGNPQASIAQIITYVVLGVIGAIYGLKNDST---- 421
415 ATODRGILYFVGAPTYGMLNAVLFVYLAVSQESQDGLYQKQOMMLAYAL-HVLP 473
422 GIONRAGVLE-FLTNOCSSVSVALEFVEKRLIHEITISGYRVSFYGLKLSLDLP 480
474 FSVVATMIFSSVCYWTGLGHPVARGFSALLAPHLIGFEFLTVLGLIYONPNIVNSY 533
481 MMLPSIIFTCIYFIMLGKPKADAFVMMFTLM--WVAYSASSMALAIAGOSVSVSA 537
534 VALLSIAGY--LVGSGFLNIQIEMPIPEKTIISYTFQKCYSEILVYNEFYGLNFTCGSSN 591
538 TLMATICVFYFMIFSGLLVNLTTIASWLSMLOFYFISIPRYGFTALQHNFEIAGNFCPG--- 594
592 VSVTNPMCAFTQGIQIEKTCPG 615
595 LNATGNPCNTA-----TCTG 610

RESULT 11
US-10-120-687-61
Sequence 61, Application US/10120687
Publication No. US20030082155A1
GENERAL INFORMATION:
APPLICANT: Massachusetts General Hospital
TITLE OF INVENTION: Stem Cells of the Islets of Langerhans and Their Use in Treating Diabetes Mellitus
FILE REFERENCE: 3284/12358
CURRENT APPLICATION NUMBER: US/10/120,687
PRIORITY FILING DATE: 2002-04-11
PRIOR APPLICATION NUMBER: US60/169082
PRIORITY FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: US 09/963,875
PRIORITY FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/215109
PRIORITY FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: US 60/238880

Wed Jun 11 09:51:14 2003

us-09-989-981a-6.rapb

Page 7

PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: US 09/731261
PRIOR FILING DATE: 2000-12-06
NUMBER OF SEQ ID NOS: 61
SOFTWARE: PatentIn version 3.1
SEQ ID NO 61
LENGTH: 655
TYPE: prt
ORGANISM: Homo sapiens
US-10-120-687-61

Query Match 20.5%: Score 680.5; DB 9; Length 655;
Best Local Similarity 29.2%; Pred. No. 2,7e-52;
Matches 182; Conservative 137; Mismatches 250; Indels 55; Gaps 18;

21 SSSLEGAPATAP---EPHSIGILHASTSVSHRVRPMDITSCROOITROIKDVSLEYE 77
13 SSGNTNGPPTASNDLKAFTEGAVLSPFNICRYVKGFLPCRPVEKEILSNINGIMK 72
78 SGQIMCIGSSGSGCTTLLDAMSGRLGAGTFLGEVYNGRALARREPODCEFSYVLSDT 137
73 PG-LNALIGPTGGGSSLLDVLAAKDPSC--LSGDVLING--APRANFKCNGSVYVDDV 129
138 LLSLTVRETLHYTALLAIRGNG--SPQKVEANMALSLSHVADRLIGNTSLGISTG 196
130 VMGTLTVRENTQFSALALATTTNHEKERNINRYTOELGLDKVADSRTGQPIRGVSG 189
197 ERRRVSAOILLDOPKVALFDEPTTGLDCAANOIIVLVLELARNRIIVLTIHOPREL 256
190 ERKRTSIGEMELITPSTILFDEPTTGLDSTANAVLLKRMKSGKRTIITSIHOPRISI 249
257 FOLDKTALISFGLIFGCTPAEMLDFFNDCGYPCPEHSNPFDEYMDLTSVDTQ---SK 312
250 FKLPDSITLLASGRIMFHPAQEALGYFESAGYHCEAVNPNADPELDIINDSTAVALNR 309
313 ERE-----IETSR-----VOMISAYKKSALCHKT-----LNIEEMKHLKTLPMYPF 356
310 EEDKATEIIPSPQDPLIEKLAETIYNSFYETAELHQLSGEKKKRTIYFEKEISY 369
357 KTKDSPGVSKLGLVLRVTNLRNKLAVITRLLQNLIMGLFL--LEFVLRVRSNVLK 414
370 TT-----SFCHQLRWKRSRPFKNLGNQASIAQIIVYVGLVIGAITFGKLNST----- 421
415 AIDRVGLIYOFVATPTTGLNANVLEPVLRAVSDESODGLYOKQOMLAVAL-HYLP 473
422 GIONRAVLF-FLTTNOCFSSVSAVELFVEKKLFIEHYISGYRVSFYFGKLLSDLP 480
474 FSVATMIFSSVCWTGLHPEVARPGYFSALLAPHLIGELTLVLGLIVQNPNIYNSV 533
481 MRMLPSTIFTCIYFEMGLKPKADAFVMMFTLM---MVAYSASSMALATAAGSVSVA 537
534 VALLSINGV--LVSGFLRNIOEMPIPKIISYFTPOKCYSEILVNEFEGLNFTCCSSN 591
538 TLMATICEFVMMIFSSGLVNLITIASLSMLQITFIRYGTALQHNHEFLQNCPCP--- 594
592 VSVTTPMCAFTGCIQIETKPCG 615
595 LMATGNPCMYA-----TCYG 610

RESULT 12
US-10-090-455-5

Sequence 5, Application US/10090455
Publication No. US20030027259A1
GENERAL INFORMATION:
APPLICANT: Chen, Hongyun
TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
FILE REFERENCE: 100103.406
CURRENT APPLICATION NUMBER: US/10/090,455
CURRENT FILING DATE: 2002-03-01
NUMBER OF SEQ ID NOS: 17
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 5
LENGTH: 655
TYPE: prt
ORGANISM: Homo sapiens
US-10-090-455-5

Query Match 20.3%: Score 674.5; DB 9; Length 655;
Best Local Similarity 29.0%; Pred. No. 9,5e-52;
Matches 181; Conservative 137; Mismatches 251; Indels 55; Gaps 18;

21 SSSLEGAPATAP---EPHSIGILHASTSVSHRVRPMDITSCROOITROIKDVSLEYE 77
13 SSGNTNGPPTASNDLKAFTEGAVLSPFNICRYVKGFLPCRPVEKEILSNINGIMK 72
78 SGQIMCIGSSGSGCTTLLDAMSGRLGAGTFLGEVYNGRALARREPODCEFSYVLSDT 137
73 PG-LNALIGPTGGGSSLLDVLAAKDPSC--LSGDVLING--APRANFKCNGSVYVDDV 129
138 LLSLTVRETLHYTALLAIRGNG--SPQKVEANMALSLSHVADRLIGNTSLGISTG 196
130 VMGTLTVRENTQFSALALATTTNHEKERNINRYTOELGLDKVADSRTGQPIRGVSG 189
197 ERRRVSAOILLDOPKVALFDEPTTGLDCAANOIIVLVLELARNRIIVLTIHOPREL 256
190 ERKRTSIGEMELITPSTILFDEPTTGLDSTANAVLLKRMKSGKRTIITSIHOPRISI 249
257 FOLDKTALISFGLIFGCTPAEMLDFFNDCGYPCPEHSNPFDEYMDLTSVDTQ---SK 312
250 FKLPDSITLLASGRIMFHPAQEALGYFESAGYHCEAVNPNADPELDIINDSTAVALNR 309
313 ERE-----IETSR-----VOMISAYKKSALCHKT-----LNIEEMKHLKTLPMYPF 356
310 EEDKATEIIPSPQDPLIEKLAETIYNSFYETAELHQLSGEKKKRTIYFEKEISY 369
357 KTKDSPGVSKLGLVLRVTNLRNKLAVITRLLQNLIMGLFL--LEFVLRVRSNVLK 414
370 TT-----SFCHQLRWKRSRPFKNLGNQASIAQIIVYVGLVIGAITFGKLNST----- 421
415 AIDRVGLIYOFVATPTTGLNANVLEPVLRAVSDESODGLYOKQOMLAVAL-HYLP 473
422 GIONRAVLF-FLTTNOCFSSVSAVELFVEKKLFIEHYISGYRVSFYFGKLLSDLP 480
474 FSVATMIFSSVCWTGLHPEVARPGYFSALLAPHLIGELTLVLGLIVQNPNIYNSV 533
481 MRMLPSTIFTCIYFEMGLKPKADAFVMMFTLM---MVAYSASSMALATAAGSVSVA 537
534 VALLSINGV--LVSGFLRNIOEMPIPKIISYFTPOKCYSEILVNEFEGLNFTCCSSN 591
538 TLMATICEFVMMIFSSGLVNLITIASLSMLQITFIRYGTALQHNHEFLQNCPCP--- 594
592 VSVTTPMCAFTGCIQIETKPCG 615
595 LMATGNPCMYA-----TCYG 610

RESULT 13

US-09-866-866A-10
Sequence 10, Application US/09866866A
Patent No. US20020102244A1
GENERAL INFORMATION:
APPLICANT: Sorrentino, Brian
TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
FILE REFERENCE: 1340-1-021CIP2
CURRENT APPLICATION NUMBER: US/09/866,866A
CURRENT FILING DATE: 2001-08-30
PRIOR APPLICATION NUMBER: 09/584,586
PRIOR FILING DATE: 2000-05-31
PRIOR APPLICATION NUMBER: PCT/US99/11825
PRIOR FILING DATE: 1999-05-27
PRIOR APPLICATION NUMBER: 60/086,988
PRIOR FILING DATE: 1998-05-28
NUMBER OF SEQ ID NOS: 27

```

: NUMBER OF SEQ ID NOS: 27
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 27
: LENGTH: 655
: TYPE: prt
: ORGANISM: Homo sapien
us-09-866-866A-27

Query Match      20.2%; Score 672.5; DB 10; Length 655;
Best Local Similarity 29.0%; Pred. No.1.4e-51;
Matches 181; Conservative 137; Mismatches 251; Indels 55; Gaps 18;

QY      21  SSSLEGGAPAPAP---EPHSLGILHASSVSHRRAPRPMMDITSCROQRTROLKQVSLAYE 77
      || : ||| | | | | | : : : : : : : : : : ||| : ||| : : :
Db      13  SGGNTNGPPAPASNDLKKFTGCAVLSFNITCYRVLKSGFLPCRRPYEKELISNNGIMK 72
      || : ||| | | | | | : : : : : : : : : : ||| : ||| : : :

QY      78  SGOIMCIIAGSSSGSGKTTLLDMSGRLGAGTFLGEVYVNGRALRREDOCPDESYAGSDT 137
      | : ||| | | | | | | | | : | : | | | | | | | | : | : | | |
Db      73  PG-LNALIGPFGGCKSLLDVLAARKDSC--LSGVDYLNG-APRANFKCMSGYVQDDY 129
      || : ||| | | | | | | | | : | : | | | | | | | | : | : | | |

QY      138  LLSLTVRETLHYTALLAIRGNPG-SFOKVEAVMAELSLSHVADRLLIGNYSLGISTG 196
      || : ||| | | | | | : : : : : | : | | | | | | | | : | : | | |
Db      130  VMGLTVRENIQFSALELATTTMTNHEKNERINRVIQELGDKVADKSVGQFINGVSGG 189
      || : ||| | | | | | : : : : : | : | | | | | | | | : | : | | |

QY      197  ERRRVSTAAQLQDPKVMLEPPTTGLDCMTANOIVLVVLVLAARRNRIVLTIHOPREL 256
      || : ||| | | | | | | | | : | : | | | | | | | | : | : | | |
Db      190  ERKRTSIGEMELITPESILFDEPPTGLDSSTANAVLLIKRMKSGRTIIFSIHQPRYSI 249
      || : ||| | | | | | | | | : | : | | | | | | | | : | : | | |

QY      257  FQLEPKIAISFEGELIFCGTPAEMLDPENDGYPCEPHSNPFDPMOLTSVDTO----SK 312
      || : ||| | | | | | | | | : | : | | | | | | | | : | : | | |
Db      250  FKLDLSLTLASGLMRHGPQOELGYFESAGYHCEAVNNADPFLDITINDSTAVALNR 309
      || : ||| | | | | | | | | : | : | | | | | | | | : | : | | |

QY      313  ERE-----IETSKR---VOMISAVYKKAICHTK-----LKNIEBKHLKTLPMYVF 356
      || : ||| | | | | | : : : : : | : | : | | | | | | : | : | | |
Db      310  EEDFKATEIIEPSSKQDKPLIEKLAIEYVNSFYETKAEHLQSGEKKKIITYFKELISY 369
      || : ||| | | | | | | | | : | : | | | | | | | | : | : | | |

QY      357  KTKSPGVFSLGVLKLRVRYTNLVKRLKAVITRLLONLIMGLFL--LEFVLKPSNWLKG 414
      || : ||| | | | | | | | | : | : | | | | | | | | : | : | | |
Db      370  TT-----SFCHQLRWVSKRSPFNLLGNPDASIAQIIVVVLGLVIGAIYFGKLNJST---- 421
      || : ||| | | | | | | | | : | : | | | | | | | | : | : | | |

QY      415  AIDRVGLLYOFVATPYTGMILANVNLFPVLRAVSDOSQGLYOKMOQMLAYAL-HVLP 473
      || : ||| | | | | | | | | : | : | | | | | | | | : | : | | |
Db      422  GIORAGVLF-FLTTNOCFSSVSNAVELFVYEKKLFIHEYISGYIRVSSYFLGKLSLDLP 480
      || : ||| | | | | | | | | : | : | | | | | | | | : | : | | |

QY      474  FSVAVTAMFSSVCYWTGLGLHDEVARFGYFSALLAPHLIGELTVLVLGIYQNPINYSV 533
      || : ||| | | | | | | | | : | : | | | | | | | | : | : | | |
Db      481  MRMLPSIIFTCIYVFMGLKKAKADAFVMMFTLM---MVAVSASMSALAIANAAGSVSVA 537
      || : ||| | | | | | | | | : | : | | | | | | | | : | : | | |

QY      534  VALLSINGV--LVSSGFLRNQEMPIPKIISYTFQPCGEIIVNVEFTLNTCGGSN 591
      || : ||| | | | | | | | | : | : | | | | | | | | : | : | | |
Db      538  TLLMTTGFVFMIMPSGLVNLTLTASMLSMLOYFISIPRYGPTALQHNHFLGQINCPG--- 594
      || : ||| | | | | | | | | : | : | | | | | | | | : | : | | |

QY      592  VSVTTPMCAFTQGIQFLEKTCPG 615
      || : ||| | | | | | : : : : : | : | : | | | | | | : | : | | |
Db      595  LMTGNNPCNVA-----TCTG 610
      || : ||| | | | | | : : : : : | : | : | | | | | | : | : | | |

RESULT 15
US-09-866-866A-14
: Sequence 14, Application us/09866866A
: Patent No. US20020102244A1
: GENERAL INFORMATION:
: APPLICANT: Sorrentino, Brian
: APPLICANT: Schuetz, John
: TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
: FILE REFERENCE: 1340-1-021C12
: CURRENT APPLICATION NUMBER: us/09/866,866A
: PRIOR APPLICATION NUMBER: 09/354,586
: PRIOR FILING DATE: 2000-05-31
: PRIOR APPLICATION NUMBER: PCT/US99/11825
: PRIOR FILING DATE: 1999-05-27
: PRIOR APPLICATION NUMBER: 60/086,988

```

PRIOR FILING DATE: 1998-05-28
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn version 3.0
SEQ ID NO 14
LENGTH: 657
TYPE: PRT
ORGANISM: Mus musculus
US-09-866-866A-14

Query Match 19.8%; Score 660; DB 10; Length 657;
Best Local Similarity 28.0%; Pred No. 1.9e-50;
Matches 181; Conservative 135; Mismatches 242; Indels 88; Gaps 19;

```
OY 13 MGLQVNGSQSSLEGAPATAPPHSLGILHASYSVSHRVRPMWDITSCROQWTRQILKDY 72
DB 12 MSQRNNNGLPFRNNSRAVRLTLAGDVLSPFHITRYV--KKSGLV---KTYEKELSDI 66
OY 73 SLXVESGOIMCILLSSGSGKTTLLDAMSGRLRAGTGLGEVYVNGRALRREQFODCSYV 132
DB 67 NQIMKPG-LMAILGPTGGKSSLDVLAARKDPKG-LSGDVLNG-APQPAHFKCCSGYV 123
OY 133 LQSDTLLSLVRETLHYTALALAIRGNPSFO-----KKVAVNAELSLSHVADRLIGN 187
DB 124 VQDDVVMGLTYREMLQPSALRL---PTMKNHKNERINTIIEIGLEKRVADSKYGT 179
OY 188 YSLGISTGERRRVSIAQLQDPKVMLEDEPTGLDCTANQIVLVLELARNRRIYVL 247
DB 180 QFIRGISGGERKRTSIGMELITDPSILFLEDEPTGLDSTANAVILLKRMKQKRTIIF 239
OY 248 TIHORSEIFOLFDIALISGELIFCGTPAPMLDEFNDGCGPCPEHNSNPFDMYDLTSV 307
DB 240 SIHORYSIFKLFDSITLILASGKLVFHPAKALEFPASAGYHCPYNNPADFLDYING 299
OY 308 PTOS-----KEREFTSR-----VOMIESAYKKSALCHTKLNIERMKHUKTLP 352
DB 300 DSSAVMLNREDDNEANKTEEPSKGEKPIYENLSEFYINSALYGETKAEILDQ----- 352
OY 353 MPPFTKDSPGVFSKLG-----LRLRYTNLYRNKLAIVITRLQNL 394
DB 353 -----PGAQEKKTSAFKEPYVTSFCHQLRIARSKKNLGNPQASVAQILIVT 403
OY 395 IMGLFL--LFTVLAVRSNVLKGAIDRVGLYQFVGATPYTGMNAVNLFPVLRASDOE 452
DB 404 ILGLITGAIFYDLKYDA---AGMQNRAGVLF-FLTNQCFSSVASVELFVEYKELFHE 458
OY 453 SODGIYOKROMLAVAL-HVLPFSVATMISSVCYTWTGLHPEVARFGYSALLAPHL 511
DB 459 YISGIYRVSSYFPGKVMSDLPMRFLPSVIFTCLYFMGLKTYDAFIMFTLI---M 515
OY 512 IGEFTLLVLLGIYQNPRIVNSVALLIAGV--LVGSGFLRNIOEMIPFKIISYTRQK 569
DB 516 VAYTASSMALAIAGOSVSVATILMTIAFVPMLEFSLVNLKRTIGPMLSMLOYESTPR 575
OY 570 YCSEILVNEFYGLNFTCGSSNVSTYTNPMCAFTOGIOFIKTCPG 615
DB 576 YGFTALQYNEFLGOEFCPG---FNVTDNSTCVNSYAL-----CTG 612
```

Search completed: June 11, 2003, 09:14:46
Job time : 26 secs

Page 1

OM nucleic - nucleic search, using sw model

(Without alignments)
9675.628 Million cell updates/sec

Title:	US-09-989-981A-5
Perfect score:	3240

Sequence: 1 gtcaggtgagcagcagg,.....aatatcataaacctatgg 2340

Scoring table: IDENTITY_NUC

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 08

Maximum Match 1008
Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubnpa/US07_PUBCOMB.seq: *
- 2: /cgn2_6/ptodata/2/pubnpa/PCT_NEW_PUB.seq: *
- 3: /cgn2_6/ptodata/2/pubnpa/US06_NEW_PUB.seq: *
- 4: /cgn2_6/ptodata/2/pubnpa/US06_PUBCOMB.seq: *
- 5: /cgn2_6/ptodata/2/pubnpa/US07_NEW_PUB.seq: *
- 6: /cgn2_6/ptodata/2/pubnpa/PCUS_PUBCOMB.seq: *
- 7: /cgn2_6/ptodata/2/pubnpa/US08_NEW_PUB.seq: *
- 8: /cgn2_6/ptodata/2/pubnpa/US08_PUBCOMB.seq: *
- 9: /cgn2_6/ptodata/2/pubnpa/US09_PUBCOMB.seq: *
- 10: /cgn2_6/ptodata/2/pubnpa/US09_PUBCOMB.seq: *
- 11: /cgn2_6/ptodata/2/pubnpa/US09_PUBCOMB.seq: *
- 12: /cgn2_6/ptodata/2/pubnpa/US09_NEW_PUB.seq: *
- 13: /cgn2_6/ptodata/2/pubnpa/US10_PUBCOMB.seq: *
- 14: /cgn2_6/ptodata/2/pubnpa/US10_PUBCOMB.seq: *
- 15: /cgn2_6/ptodata/2/pubnpa/US10_NEW_PUB.seq: *
- 16: /cgn2_6/ptodata/2/pubnpa/US10_NEW_PUB.seq: *
- 17: /cgn2_6/ptodata/2/pubnpa/US60_PUBCOMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2340	100.0	2340	9	US-09-837-992-4	Sequence 4, Appl
2	2340	100.0	2340	11	US-09-989-981A-5	Sequence 5, Appl
3	1395.6	59.6	2258	9	US-09-837-992-2	Sequence 2, Appl
4	1365.4	58.4	1959	11	US-09-989-981A-1	Sequence 1, Appl
5	472	20.6	472	9	US-09-837-992-19	Sequence 19, Appl
6	249	10.6	249	9	US-09-837-992-7	Sequence 7, Appl
7	214	9.1	214	9	US-09-837-992-14	Sequence 14, Appl
8	206	8.8	206	9	US-09-837-992-15	Sequence 15, Appl
9	203.6	8.7	2669	11	US-09-989-981A-7	Sequence 7, Appl
10	199.2	8.5	2019	11	US-09-989-981A-3	Sequence 3, Appl
11	186	7.9	186	9	US-09-837-992-17	Sequence 17, Appl
12	140	6.0	140	9	US-09-837-992-12	Sequence 12, Appl
13	139.6	6.0	759	13	US-10-027-632-152155	Sequence 152155
14	139.6	6.0	759	13	US-10-027-632-152156	Sequence 152156
15	139.6	6.0	759	13	US-10-027-632-152157	Sequence 152157
16	139	5.9	139	9	US-09-837-992-16	Sequence 16, Appl

ALIGNMENTS

17	137	5.9	137	9	US-09-837-992-9	Sequence 9, Appl1
18	135.4	5.8	472	11	US-09-918-895-30637	Sequence 30637, A
19	130	5.6	130	9	US-09-837-992-13	Sequence 13, Appl1
20	129	5.5	129	9	US-09-837-992-11	Sequence 11, Appl1
21	122	5.2	122	9	US-09-837-992-8	Sequence 8, Appl1
22	114.2	4.9	2930	10	US-09-554-531-591	Sequence 591, Appl1
23	114.2	4.9	2930	10	US-10-171-581-276	Sequence 276, Appl1
24	114.2	4.9	3201	13	US-10-072-621-5	Sequence 5, Appl1
25	113	4.8	113	9	US-09-837-992-18	Sequence 18, Appl1
26	110.8	4.7	2687	13	US-10-154-952-3	Sequence 3, Appl1
27	110.8	4.7	2687	14	US-10-090-455-12	Sequence 12, Appl1
28	107.6	4.6	2687	13	US-10-154-452-7	Sequence 7, Appl1
29	107.2	4.6	1941	14	US-10-090-455-3	Sequence 3, Appl1
30	107.2	4.6	3455	13	US-10-072-621-4	Sequence 4, Appl1
31	107.2	4.6	3455	14	US-10-090-455-1	Sequence 1, Appl1
32	103	4.4	103	9	US-09-837-992-10	Sequence 10, Appl1
33	101.6	4.4	6043	11	US-09-889-981A-9	Sequence 9, Appl1
34	98	4.2	2247	10	US-09-866-866A-26	Sequence 26, Appl1
35	98	4.2	2400	13	US-10-108-605-244	Sequence 244, Appl1
36	98	4.2	2418	11	US-09-861-086-2	Sequence 2, Appl1
37	98	4.2	2574	10	US-09-881-553-34	Sequence 34, Appl1
38	98	4.2	2718	14	US-10-120-867-9	Sequence 60, Appl1
39	98	4.2	2719	10	US-09-866-866A-9	Sequence 9, Appl1
40	96.4	4.1	10330	13	US-10-001-189-68	Sequence 68, Appl1
41	90.4	3.9	2025	10	US-09-866-866A-13	Sequence 13, Appl1
42	88.4	3.8	2223	10	US-09-938-848A-2262	Sequence 2262, Appl1
43	83	3.5	3376	14	US-10-037-270-918	Sequence 918, Appl1
44	82.4	3.5	2788	9	US-09-745-763-316	Sequence 136, Appl1
45	72.8	3.1	427	10	US-09-660-352-12839	Sequence 12839, A

US-09-837-992-4
 RESULT 1
 Sequence 4, Application US/09837992
 Patent No. US20020081687A1
 GENERAL INFORMATION:
 APPLICANT: Tian, Hui
 APPLICANT: Schultz, Joshua
 APPLICANT: Shan, Bei
 APPLICANT: Tularik Inc.
 TITLE OF INVENTION: Sltosterolemia Susceptibility Gene (SSG): Compositions
 TITLE OF INVENTION: and Methods of Use
 FILE REFERENCE: 018781-00602005
 CURRENT APPLICATION NUMBER: US/09/837,992
 CURRENT FILING DATE: 2001-04-18
 PRIOR APPLICATION NUMBER: US 60/198,465
 PRIOR FILING DATE: 2000-04-18
 PRIOR APPLICATION NUMBER: US 60/204,234
 PRIOR FILING DATE: 2000-05-15
 NUMBER OF SEQ ID NOS: 45
 SOFTWARE: Patentln Ver. 2.1
 SEQ ID NO 4
 LENGTH: 2340
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: human sltosterolemia gene (SSG)
 NAME/KEY: CDS
 LOCATION: (107)..(2062)
 OTHER INFORMATION: human sltosterolemia susceptibility gene (SSG)
 OTHER INFORMATION: protein
 US-09-837-992-4

	Query Match	100.0%	Score 2340;	DB 9;	Length 2340;
	Best Local Similarity	100.0%;	Pred. No. 0;		
	Matches 2340;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
Qy	1	GTGAGGTGAGACGAGCGAGGCGAGTCTGCCACGCGCTCCCAACTGATGAGCCACTCTGGGGA	60		
db	1	GTGAGGTGAGACGAGCGAGGCGAGTCTGCCACGCGCTCCCAACTGATGAGCCACTCTGGGGA	60		

OY 61 GGGTCGGGACACGAAATTTGGCCAGCTTTGGCTGTGGCCATGGGTGACCTCTC 120
| | | | |
Db 61 GGGTCGGGACACGAAATTTGGCCAGCTTTGGCTGTGGCCATGGGTGACCTCTC 120
OY 121 ATCTTTAACCCCGGGAGGCTCAGTGGTCTCCAGATTAACAGAGGCTCCAGAGCTCCT 180
| | | | |
Db 121 ATCTTTAACCCCGGGAGGCTCAGTGGTCTCCAGATTAACAGAGGCTCCAGAGCTCCT 180
OY 181 GAGAGGGGGCTCTGGCCACGGCCCGGGAGCTTCACAGCTGGGGATCCTCCATGCTCTCTA 240
| | | | |
Db 181 GAGAGGGGGCTCTGGCCACGGCCCGGGAGCTTCACAGCTGGGGATCCTCCATGCTCTCTA 240
OY 241 CAGGCTACGCCACCGCTGTAGGCGCTGTGGGACATCATCTTCCGCGGACAGTGGAC 300
| | | | |
Db 241 CAGGCTACGCCACCGCTGTAGGCGCTGTGGGACATCATCTTCCGCGGACAGTGGAC 300
OY 301 CAGGAGATCTCTCAAGATGTCTCTTGTACGTGGAGGCGGAGATCATGTGCATCT 360
| | | | |
Db 301 CAGGAGATCTCTCAAGATGTCTCTTGTACGTGGAGGCGGAGATCATGTGCATCT 360
OY 361 AGGAGCTCAGGCTCGGGGAAACACAGCTGTGGAGCGGATGTCCGGGAGGCTGGGCG 420
| | | | |
Db 361 AGGAGCTCAGGCTCGGGGAAACACAGCTGTGGAGCGGATGTCCGGGAGGCTGGGCG 420
OY 421 CCGGGGAGCTTCTGTGGGAGGTGTATGTGAACGGCGCGGCGCTGCGCGGAGGAGT 480
| | | | |
Db 421 CCGGGGAGCTTCTGTGGGAGGTGTATGTGAACGGCGCGGCGCTGCGCGGAGGAGT 480
OY 481 CCGAGACTGCTTCTCTAGCTCTGTGACAGACGACCTGTGGAGAGGCTTCACCGTGGC 540
| | | | |
Db 481 CCGAGACTGCTTCTCTAGCTCTGTGACAGACGACCTGTGGAGAGGCTTCACCGTGGC 540
OY 541 CGAGAGCTGACCTACACCGCGCTGTGCGCATCCCGCGGGAATCCCGGCTCTTCCA 600
| | | | |
Db 541 CGAGAGCTGACCTACACCGCGCTGTGCGCATCCCGCGGGAATCCCGGCTCTTCCA 600
OY 601 GAAGAAGTGAAGGCGGTGATGAGCAGAGTGAATGAGCAGATGTGGCAGAGCAGCTGAT 660
| | | | |
Db 601 GAAGAAGTGAAGGCGGTGATGAGCAGAGTGAATGAGCAGATGTGGCAGAGCAGCTGAT 660
OY 661 TGGCAACTACAGTTTGGGGGCAATTTCCACAGGGTGAAGCGCGGGGTCTCCATGCGACG 720
| | | | |
Db 661 TGGCAACTACAGTTTGGGGGCAATTTCCACAGGGTGAAGCGCGGGGTCTCCATGCGACG 720
OY 721 CCAGCTGTCCAGAGATCTAAGGTGATGCTGTTGTGATGAGCCACACAGGCTGTGACG 780
| | | | |
Db 721 CCAGCTGTCCAGAGATCTAAGGTGATGCTGTTGTGATGAGCCACACAGGCTGTGACG 780
OY 781 CATGACTGCTAATCAGATTTGCTCTCTGTTGTAAGTGGCTGCGAGAGACGAAATGT 840
| | | | |
Db 781 CATGACTGCTAATCAGATTTGCTCTCTGTTGTAAGTGGCTGCGAGAGACGAAATGT 840
OY 841 GGTTCACACATTCACACAGCCCGTGTGAGCTTTTCAAGCTTTTGAACAAATTTGCCAT 900
| | | | |
Db 841 GGTTCACACATTCACACAGCCCGTGTGAGCTTTTCAAGCTTTTGAACAAATTTGCCAT 900
OY 901 CCTGAGCTTCGGAGAGCTAATTTCTGTGGCAGCGCAGGGAATGCTGATTTCTTCA 960
| | | | |
Db 901 CCTGAGCTTCGGAGAGCTAATTTCTGTGGCAGCGCAGGGAATGCTGATTTCTTCA 960
OY 961 TGAATGCGGTACCTTGTCTGAACATTCAAACCTTTTGAATCTTATATGAGCTGAC 1020
| | | | |
Db 961 TGAATGCGGTACCTTGTCTGAACATTCAAACCTTTTGAATCTTATATGAGCTGAC 1020
OY 1021 GTACAGTGAATCCCAAGAGAGGAAATAGAACTTCAAGAGAGTCCAGATGAT 1080
| | | | |
Db 1021 GTACAGTGAATCCCAAGAGAGGAAATAGAACTTCAAGAGAGTCCAGATGAT 1080
OY 1081 AGAATCTGCTTCAAGAAATCAGAAATTTGTATAAACTTTGAAGAAATTTGAAGAT 1140
| | | | |
Db 1081 AGAATCTGCTTCAAGAAATCAGAAATTTGTATAAACTTTGAAGAAATTTGAAGAT 1140

OY 1141 GAAACACCTGAAAAAGTTTACCAATGTTCTTTCAAACCAAGATTTCTCTGAGTTT 1200
| | | | |
Db 1141 GAAACACCTGAAAAAGTTTACCAATGTTCTTTCAAACCAAGATTTCTCTGAGTTT 1200
OY 1201 CTCTAAATCGGATGTTCTCTGAGAGAGTGAACAAATTTGGTGAATATAGCTGGC 1260
| | | | |
Db 1201 CTCTAAATCGGATGTTCTCTGAGAGAGTGAACAAATTTGGTGAATATAGCTGGC 1260
OY 1261 AGTGAATACGCGTCTCTTGAATCTGATCATGAGTTTGTCTCTCTTCTCTCTCT 1320
| | | | |
Db 1261 AGTGAATACGCGTCTCTTGAATCTGATCATGAGTTTGTCTCTCTCTCTCTCTCT 1320
OY 1321 GCGGGTCCGAAAGAAATGTCTAAAGGTTGCTATTCAGAGACCGGATGATCTCTTACCA 1380
| | | | |
Db 1321 GCGGGTCCGAAAGAAATGTCTAAAGGTTGCTATTCAGAGACCGGATGATCTCTTACCA 1380
OY 1381 GTTGTGGGCGCCACCCGCTTACACAGGATGTAAGCGTGTAAATCTGTTCCCGGCT 1440
| | | | |
Db 1381 GTTGTGGGCGCCACCCGCTTACACAGGATGTAAGCGTGTAAATCTGTTCCCGGCT 1440
OY 1441 GCGAGCTGTACGCGACAGAGAGTACAGAGCGGCTCTTACAGAACTGGCAGATGATCT 1500
| | | | |
Db 1441 GCGAGCTGTACGCGACAGAGAGTACAGAGCGGCTCTTACAGAACTGGCAGATGATCT 1500
OY 1501 GGGCTATGACATGACAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1560
| | | | |
Db 1501 GGGCTATGACATGACAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1560
OY 1561 GTGCTACTGAGACGCTGGGCTTACATCTTGAAGTTGGCCGATTTGAATTTTCTGCTGC 1620
| | | | |
Db 1561 GTGCTACTGAGACGCTGGGCTTACATCTTGAAGTTGGCCGATTTGAATTTTCTGCTGC 1620
OY 1621 TCTCTTGGCCCCCACTTAATTTGTAATTTCTAATCTTGTGCTACTGTGTAATGCTCA 1680
| | | | |
Db 1621 TCTCTTGGCCCCCACTTAATTTGTAATTTCTAATCTTGTGCTACTGTGTAATGCTCA 1680
OY 1681 AAATCAAAATTAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1740
| | | | |
Db 1681 AAATCAAAATTAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1740
OY 1741 ATCTGATTTCTCAGAAATATACAGAAATGCGCATCTTCTTAAATCATCATGATTT 1800
| | | | |
Db 1741 ATCTGATTTCTCAGAAATATACAGAAATGCGCATCTTCTTAAATCATCATGATTT 1800
OY 1801 TACATTCAAAAATATGAGAGATGATTTGTGATGATGATGATGATGATGATGATGAT 1860
| | | | |
Db 1801 TACATTCAAAAATATGAGAGATGATTTGTGATGATGATGATGATGATGATGATGAT 1860
OY 1861 CACTGTGACAGCTCAAAATGTTCTGTGACAACTATTCAAATGTGCTTACTGACG 1920
| | | | |
Db 1861 CACTGTGACAGCTCAAAATGTTCTGTGACAACTATTCAAATGTGCTTACTGACG 1920
OY 1921 AATTCATTCATTTGAAGAAACCTGCCAGAGTGCACATCATGATTCACATGAACTTCT 1980
| | | | |
Db 1921 AATTCATTCATTTGAAGAAACCTGCCAGAGTGCACATCATGATTCACATGAACTTCT 1980
OY 1981 GAATTTGTATTTATTTATTCAGCTCTTGTCAATCTGAGAAATGTTGTTTCAAAATTAAG 2040
| | | | |
Db 1981 GAATTTGTATTTATTTATTCAGCTCTTGTCAATCTGAGAAATGTTGTTTCAAAATTAAG 2040
OY 2041 GGAATCATCTATAGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2100
| | | | |
Db 2041 GGAATCATCTATAGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2100
OY 2101 ACTGTGATGATGCTGTGAAAGTGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2160
| | | | |
Db 2101 ACTGTGATGATGCTGTGAAAGTGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2160
OY 2161 GACATCTCAAGTCTTTTAAACATTTAAGACTCATTTGTGCTCTTGGATCCAGAGGCG 2220
| | | | |
Db 2161 GACATCTCAAGTCTTTTAAACATTTAAGACTCATTTGTGCTCTTGGATCCAGAGGCG 2220
OY 2221 TTGAATGCAATGAGAGTGTATATGATGCTTGTCTTACAACTTGCAGAGGAGATGTGT 2280
| | | | |

Db 2221 TTGATCCAAATGAGAGGTTTATAGTCCCTGCTTACAACTTGACAGGACATGTGT 2280
QY 2281 TATTGGAAATGTCAGTCAGACGACCAAGAAATGTAATTAATTCATAAAGCTATAGG 2340
Db 2281 TATTGGAAATGTCAGTCAGACGACCAAGAAATGTAATTAATTCATAAAGCTATAGG 2340

RESULT 2

US-09-989-981A-5
: Sequence 5, Application US/09989981A
: Publication No. US20030049730A1
: GENERAL INFORMATION:
: APPLICANT: Hobbs, Helen H.
: APPLICANT: Shan, Bel
: APPLICANT: Barnes, Robert
: APPLICANT: Tian, Hui
: APPLICANT: Tularik Inc.
: APPLICANT: Board of Regents, The University of Texas System
: TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
: FILE REFERENCE: 018781-007320US
: CURRENT APPLICATION NUMBER: US/09/989, 981A
: CURRENT FILING DATE: 2002-07-23
: PRIOR APPLICATION NUMBER: US 60/252,235
: PRIOR FILING DATE: 2000-11-20
: PRIOR APPLICATION NUMBER: US 60/253,645
: PRIOR FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 13
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 5
: LENGTH: 2340
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (107)..(2062)
: OTHER INFORMATION: human ABCG5 (hABCG5)
US-09-989-981A-5

Query Match Best Local Similarity 100.0%; Score 2340; DB 11; Length 2340;
Matches 2340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCAGTGGAGCAGCAGGAGGAGTCTGCCACGGGCTCCCACTGAAAGCACTCTGGGA 60
Db 1 GTCAGTGGAGCAGCAGGAGGAGTCTGCCACGGGCTCCCACTGAAAGCACTCTGGGA 60
QY 61 GGGTCGGGACAGAGAAATTTGCCAGCTTGTGCTGCTGTTGGCCATGGTGACCTCTC 120
Db 61 GGGTCGGGACAGAGAAATTTGCCAGCTTGTGCTGCTGTTGGCCATGGTGACCTCTC 120
QY 121 ATCTTGAACCCCGGAGAGGTCCATGCTTCCAAAGTAACAGAGGCTCCAGAGCTCCCT 180
Db 121 ATCTTGAACCCCGGAGAGGTCCATGCTTCCAAAGTAACAGAGGCTCCAGAGCTCCCT 180
QY 121 ATCTTGAACCCCGGAGAGGTCCATGCTTCCAAAGTAACAGAGGCTCCAGAGCTCCCT 180
Db 121 ATCTTGAACCCCGGAGAGGTCCATGCTTCCAAAGTAACAGAGGCTCCAGAGCTCCCT 180
QY 181 GGAGGGGCTCTCTGACACCGCCCGGAGGCTCAGAGGCTGGGCAATCTTCATAGCTCTTA 240
Db 181 GGAGGGGCTCTCTGACACCGCCCGGAGGCTCAGAGGCTGGGCAATCTTCATAGCTCTTA 240
QY 241 CAGCTCAGACCCAGCCGCTGAGAGCCCTGTGTGAGACATCAATCTTCCGCGACAGTGGAC 300
Db 241 CAGCTCAGACCCAGCCGCTGAGAGCCCTGTGTGAGACATCAATCTTCCGCGACAGTGGAC 300
QY 301 CAGGAGATCTCTCAAGATGTCTCTTGTAGCTGAGAGCGGGCAGATCATGTGATCTCT 360
Db 301 CAGGAGATCTCTCAAGATGTCTCTTGTAGCTGAGAGCGGGCAGATCATGTGATCTCT 360
QY 361 AGGAGCTCAGGCTCCGGGAAACACAGCTGTGTGAGCGCAATGTCGGGAGAGCTGGGGG 420
Db 361 AGGAGCTCAGGCTCCGGGAAACACAGCTGTGTGAGCGCAATGTCGGGAGAGCTGGGGG 420
QY 421 CGCGGGAGCTTCTCTGGGGAGGTGTATGTGAACGGCGGCGCTGCCCGGAGAGCTT 480
Db 421 CGCGGGAGCTTCTCTGGGGAGGTGTATGTGAACGGCGGCGCTGCCCGGAGAGCTT 480

Db 421 CGCGGGAGCTTCTCTGGGGAGGTGTATGTGAACGGCGGCGCTGCCCGGAGAGCTT 480
QY 481 CCAGAGCTGCTTCTCTAGTCTCTGACAGAGCGACACCTCTGAGAGCTCAACCTGGC 540
Db 481 CCAGAGCTGCTTCTCTAGTCTCTGACAGAGCGACACCTCTGAGAGCTCAACCTGGC 540
QY 541 CGAGAGCTGCTCTACACCGCGGCTGCTGGGCAATCCGCGGCAATCCGCGCTTCCA 600
Db 541 CGAGAGCTGCTCTACACCGCGGCTGCTGGGCAATCCGCGGCAATCCGCGCTTCCA 600
QY 601 GAAGAGGTGAGAGCCCTCATGTCAGAGCTGAGCTGAGCCATGTGGCAGACAGCTGAT 660
Db 601 GAAGAGGTGAGAGCCCTCATGTCAGAGCTGAGCTGAGCCATGTGGCAGACAGCTGAT 660
QY 661 TGGCACTACAGCTTGGGGGGCATTTTCCACGGGTGAGCGGCGGGTCTCCATCCGAGC 720
Db 661 TGGCACTACAGCTTGGGGGGCATTTTCCACGGGTGAGCGGCGGGTCTCCATCCGAGC 720
QY 721 CCAGCTGCTCCAGATCTTAAGTCAATGCTGTTGATGACCAACACAGGCTTGACTG 780
Db 721 CCAGCTGCTCCAGATCTTAAGTCAATGCTGTTGATGACCAACACAGGCTTGACTG 780
QY 781 CATGACTGCTAATCAGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
Db 781 CATGACTGCTAATCAGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
QY 841 GGTTCACACATTCACACAGCCCGTCTGAGCTTTTTCAGCTTTTTCAGCTTTTTCAG 900
Db 841 GGTTCACACATTCACACAGCCCGTCTGAGCTTTTTCAGCTTTTTCAGCTTTTTCAG 900
QY 901 CCTGAGCTTGGAGAGCTGATTTTCTGTGACACGCCAGCGGAAATCTGTATTTTCA 960
Db 901 CCTGAGCTTGGAGAGCTGATTTTCTGTGACACGCCAGCGGAAATCTGTATTTTCA 960
QY 961 TGACTGCGTTTACCTTGTCTGACATTCAAACCTTTTGAATTTTATATGACCTGAC 1020
Db 961 TGACTGCGTTTACCTTGTCTGACATTCAAACCTTTTGAATTTTATATGACCTGAC 1020
QY 1021 GTCAGTGGATACCAAGCAAGGAGGAGAAATAGAAACCTCCAGAGAGTCCAGATGAT 1080
Db 1021 GTCAGTGGATACCAAGCAAGGAGGAGAAATAGAAACCTCCAGAGAGTCCAGATGAT 1080
QY 1081 AGAATCTGCTTACAGAAATTCAGAAATTTTGTATTAACCTTGAAGAAATTTGAAGAAT 1140
Db 1081 AGAATCTGCTTACAGAAATTCAGAAATTTTGTATTAACCTTGAAGAAATTTGAAGAAT 1140
QY 1141 GAAACACGTGAAGAGTTTCAATGCTTCTTCAAAACCAAGATTTCTCTGAGCTTTT 1200
Db 1141 GAAACACGTGAAGAGTTTCAATGCTTCTTCAAAACCAAGATTTCTCTGAGCTTTT 1200
QY 1201 CTCTAACTGGGTGTTCTCTGAGAGAGTGAACAAGAACTTGTGAGAAATTAAGTGGC 1260
Db 1201 CTCTAACTGGGTGTTCTCTGAGAGAGTGAACAAGAACTTGTGAGAAATTAAGTGGC 1260
QY 1261 AGTGAATACGCTCTCTCAAGATCTGATGAGTGTGTTCTCTCTCTCTCTCTCTCT 1320
Db 1261 AGTGAATACGCTCTCTCAAGATCTGATGAGTGTGTTCTCTCTCTCTCTCTCTCT 1320
QY 1321 GGGGGTCCGAAGCAATGCTTCAAGAGGTCTATCAGAGACCGGCTAGAGTCTCTTACCA 1380
Db 1321 GGGGGTCCGAAGCAATGCTTCAAGAGGTCTATCAGAGACCGGCTAGAGTCTCTTACCA 1380
QY 1381 GTTTGAGGCGCCACCCCTTACACAGAGCTGCTGAAGCGCTGTGANTCTGTTCCGCTGT 1440
Db 1381 GTTTGAGGCGCCACCCCTTACACAGAGCTGCTGAAGCGCTGTGANTCTGTTCCGCTGT 1440
QY 1441 GCGAGCTGTCAGCAACAGAGAGATGTCAGAGCGGCTTACAGAAATGGCAGATGATCT 1500
Db 1441 GCGAGCTGTCAGCAACAGAGAGATGTCAGAGCGGCTTACAGAAATGGCAGATGATCT 1500
QY 1501 GGCCTATGACATGACAGCT 1560
Db 1501 GGCCTATGACATGACAGCT 1560

Query Match	Best Local Similarity	Score	DB	Length
Matches 1642: Conservative	80.7%	1395.6	9	2258
		Pred. No. 0;		
		Mismatches 389; Indels 3; Gaps 1;		
61	GGGTCGGGGCCACGAGAAATTTGGCCAGCTTGTGCTGCTGTTGGCCATGGGTACCTCTC	120		
1	GGGACAGGCCCTGAGAAATTTCACTTCATTTCTCTCTGTACCACTGGGTGAGCTGCC	60		
121	ATCTTTGACCCCCCGAGGGTTCATGCGCTTCCAGTAACAGAGAGCTCCAGACTCCCT	180		
61	CTTTCTGTGCTCAAGAGGAGCCAGAGGGCCCTACATACAGAGAGGCTCTGAGCTCCT	120		
181	GGAGGGGCTCTCTCCACACGCCCGCGAGACT--CACAGCTGGGCACTCTCATGCTC	237		
121	GGACCAAGTTCGGTCAAGGCGCACAGAGGCTCGCACAGCTTAGGTGCTCGATGTGTC	180		
238	CTACAGGCTGACGACCGCGTGGAGCCCTGGTGGAGATCATCATCTTGGCCGACGAGCTG	297		
181	CTACAGGCTGACGACCCGTCTGCGGCTTGTGGACATCAATCATGACAGCAGAACTG	240		
298	GACCAAGCAGATCTCAAGATGTCTCTTGTACGTGGAGAGCGGAGACATCATGTGCTAT	357		
241	GGACAGGCAAAATCTCAAGATGTCTCTTGTACGTGGAGAGCGGAGACATCATGTGCTAT	300		
358	CTTGGAAAGCTCAGAGCTCCGGGAAACACAGGCTGCTGAGCCATGTCTGGAGGCTGG	417		
301	CTTAGGAGCTCAGGCTCAGAGGAGACACGCTGCTGAGCCATCTCCGGAGGCTCGC	360		
418	CGCGCGGGGACCTTCTTGGGGAGAGTATGTATGTAAGCGCGGGCGCTGCGCGGAGCA	477		
361	CGCGACTGGGACCTTGGAGGGGAGGTGTTGTAAGGCTGCGAGCTGCGAGGAGCA	420		
478	GTTCCAGAGACTGCTTCTCTTACGCTCTGACAGAGCAGACCTGCTGAGACGCTCACGCT	537		
421	GTTCCAGAGACTGCTTCTCTTACGCTCTGACAGAGCAGAGCTTCTGAGACGCTCACGCT	480		
538	CGCGAGAGCTGCTGACTACACCGCGGCTGAGGACATCGCGCGGGCAATCCCGGCTCCT	597		
481	CGCGAGAGCTTGGATACACAGCAGATGCTGGCCCTTCTGCGGAGGCTCCGGGACTCTTA	540		
598	CCAGAGAGGAGGAGGCGCTCATGCGAGAGCTGAGTCTGAGCCATGTGGACAGAGCT	657		
541	CACAGAGAGGAGGAGGAGCTCATGAGAGAGCTGAGCTGAGCCAGTGGCGGAGCCAAAT	600		
658	GATTGGCAATTAAGCTTGGGGGCAATTTCCACAGGGGAGAGCGGCGCGGCTCTCATGCG	717		
601	GATTGGCAATTAATTTTGGGGGAAATTTTCAATGGCGAGCGCGCGGACTTTCCATGCG	660		
718	AGCCAGAGCTCTCCAGAGATCTTAAGGCTCATGCTTGTGATGAGACCAACAGAGCCGGA	777		
661	AGCCAGAGCTCTCCAGAGATCTTAAGGCTCATGCTTGTGATGAGACCAACAGAGCTGGA	720		
778	CTGCAATGACTGTAATTCAGATTGTGCTCTCTGATGGAATGGCTGCGAGAACCAAT	837		
721	CTGCAATGACTGTAATTCAGATTGTGCTCTCTGATGGAATGGCTGCGAGAACCAAT	780		
838	TGAGGTTTCACCAATTCACAGGCGGCTGCTGAGCTTTTACAGCTCTTTGACAAATTCG	897		
781	TGAGGTTTCACCAATTCACAGGCGGCTGCTGAGCTTTTACAGCTCTTTGACAAATTCG	840		

```

OY 898 CATCTGAGCTTGGAGAGCTGATTTCTGTCGACCGCAGGAAATGCTGATTTCTT 957
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 841 CATCTGACTTACGAGAGTGTGTCTGTGGACCCGACAGAGATGCTGTGCTTCTT 900
OY 958 CAATGACTGCGGTTACCTTGTCTGTAACATTCACCTTTTATGATGAGCT 1017
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 901 CAATGACTGTGTACCCCTGTCTGTAACATTCACCTTTTATGATGAGCT 960
OY 1018 GAGCTGAGTATGCCAAGCAAGAAAGGGAATGAAACCTTCAAGAGAGTCAGAT 1077
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 961 GACATGAGTATGCCAAGCAAGAAAGGGAATGAAACCTTCAAGAGAGTCAGAT 1020
OY 1078 GATGAGTATGCCAAGCAAGAAATGTAATTTGTAAGATGTAAGAT 1137
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1021 GCTGAGTATGCTCTTCAAGAAATGTAATTTGTAAGATGTAAGAT 1080
OY 1138 AATGAAACACCTGAAACGTTACCAATGTTCTTCAAAACCAAGATTCCTGAGT 1197
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1081 AGCAGATGATCTGAAACCTTACCAATGTTCTTCAAAACCAAGATTCCTGAGT 1140
OY 1198 TTTCTGTAACCTGAGTGTCTCTGAGAGAGTGAAGAACTTGTGAAATAGCT 1257
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1141 GTTGGGCAACCTTGTGTCTCTGAGAGAGTGAAGAACTTATGAGAAATAGCA 1200
OY 1258 GGCAGTATGAGCTGCTCTCTGAGAGATGATGATGATGATGATGATGATGAT 1317
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1201 GGCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
OY 1318 TCTGCGGTCGGAAGAAATGCTAAAGGCTATCCAGAGAGAGAGAGAGAGAGAG 1377
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1261 TCTGCGGTCGGAAGAAATGCTAAAGGCTATCCAGAGAGAGAGAGAGAGAGAG 1320
OY 1378 CCAGTTTGTGGGCGCAACCCGCTTCAACAGAGATGTAAGCTGTGATGATGATGAT 1437
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1321 TCAGCTTGTGGGCGCAACCCGCTTCAACAGAGATGTAAGCTGTGATGATGATGAT 1380
OY 1438 GCTGCGAGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1381 GCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
OY 1498 GCTGCGAGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1441 GCTGCGAGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
OY 1558 TGTGTGCTAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1501 TGTGTGCTAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
OY 1618 TGTGTGCTAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1677
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1561 TGTGTGCTAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
OY 1678 CCAAAATCCAAATATAGTCAACAGTATGAGTGTGTCGATGAGAGAGAGAGAGAGAG 1737
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1621 CCAAAATCCAAATATAGTCAACAGTATGAGTGTGTCGATGAGAGAGAGAGAGAGAGAG 1680
OY 1738 TGTGTGCTAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1797
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1681 TGTGTGCTAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
OY 1798 TTTTCAATTCACAAAATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1857
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1741 TTTTCAATTCACAAAATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
OY 1858 TTTTCAATTCACAAAATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1917
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1801 CTTTCAATTCACAAAATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
OY 1918 AGGAATTCATTCATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1977
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1861 AGGGGTCCAGTTCATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
OY 1978 TCTGATTTTGTATTCATTCATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2037

```

```

Db 1921 CCTCATCTTATATGAGGTTTATCCAGAGCTGTGTCTATCTTACAGATGATTTTAAAGT 1980
OY 2038 AAGGATCATCTTCATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2091
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1981 CAGGAGCTACCTGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2034

RESULT 4
US-09-989-981a-1
Sequence 1, Application US/0998981A
Public Information No. US20030049730A1
GENERAL INFORMATION:
APPLICANT: Hobbs, Helen H.
APPLICANT: Shan, Bel
APPLICANT: Barnes, Robert
APPLICANT: Tlan, Hui
APPLICANT: Tularek Inc.
APPLICANT: Board of Regents, The University of Texas System
FILE REFERENCE: 018781-00732005
CURRENT FILING DATE: 2002-07-23
PRIOR APPLICATION NUMBER: US 60/252,235
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/253,645
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 1959
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1959)
OTHER INFORMATION: mouse ABCG5 (mABCG5)
US-09-989-981a-1

Query Match
Best Local Similarity 58.4%; Score 1365.4; DB 11; Length 1959;
Matches 1959; Conservative 0; Mismatches 361; Indels 3; Gaps 1;

OY 107 ATGGGAGAGCTCTCATCTTGTGACCCCGAGAGGTCATGAGTCTCAAGTAACAGAGGC 166
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ATGGGAGAGCTCTCTCTTGTGATGTCAGAGGAGCCAGAGAGGCTTCATCAACAGAGGG 60
OY 167 TCCAGAGAGCTCTCTGAGAGGAGGCTCTGTCACAGGCGCCGAGAGCT---CAGAGCTGGGC 223
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 TCTGTAGAGTCTCTCTGAGAGAGGAGTTCGTCAGAGGAGAGAGAGAGAGAGAGAGAGAGT 120
OY 224 ATCTTCATGCTCTCTACAGCTTCAAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 283
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 GTCTGAGATGTCTCTACAGGCTTCAAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 180
OY 284 TCCCGGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 343
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 TCCCGGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
OY 344 CAGATCATGTCATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 403
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 CAGATCATGTCATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
OY 404 TCCGGAGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 463
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 TCCGGAGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360
OY 464 CTGGGAGGAGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 523
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 CTGGGAGGAGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
OY 524 AGCAGAGCTCAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 583
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 421 AGCAGCTCAGTGTGGCGGAGACGTTGCGATACAGCGATGCTGGCCCTCTGCGCAGC 480
 Oy 584 AATCCGCGCTCTCCAGAGAGAGTGGAGCGGTGAGCGAGCTGAGTGTGAGCCAT 643
 Db 481 TCCGGGAGCTTTACACAAAGAGTAGAGGAGTATGACAGAGTGTAGCTGAGCCAC 540
 Oy 644 GTGGAGACGAGTGTGGGCACTACACAGCTTGGGGGCAATTTCCAGGAGTGGAGCGGC 703
 Db 541 GTGGGAGCAAAATGTGGAGCTATATTTTGGGGAAATTTCCAGTGGCAGGCGGCC 600
 Oy 704 CGGGCTCTCATTTGCGAGCCAGCTGCTCCAGATCTTAAGGTCAATGCTTTTATGAGCA 763
 Db 601 CGAGTTTCATGCGAGCCCACTCTTCAGAGACCCCAAGCATGATGCTAATGAGCA 660
 Oy 764 ACCAGAGGCTGAGCTGATGATGCTATGATGATGCTGCTCTGCTGAGAGCTGCT 823
 Db 661 ACCAGAGGCTGAGCTGATGATGCTAATGATGCTGCTCTGCTGAGAGCTGCT 720
 Oy 824 CGCAGGACGAAATGTGGTCTCAACATTCACAGCCCGTGTGAGCTTTTCAGCTC 883
 Db 721 CGCAGGACGAAATGTGTGATGTGATGATGATGATGATGATGATGATGATGATGATG 780
 Oy 884 TTTGACAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 943
 Db 781 TTTGACAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
 Oy 944 ATGCTGATTTCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1003
 Db 841 ATGCTGATTTCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
 Oy 1004 TTTCTATGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1063
 Db 901 TTTCTATGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 960
 Oy 1064 AAGAGATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1123
 Db 961 AAGAGATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020
 Oy 1124 AAGATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1183
 Db 1021 AAGATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080
 Oy 1184 GATTCCTGAGAGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1243
 Db 1081 GATTCCTGAGAGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140
 Oy 1244 GTGAGAAATTAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1303
 Db 1141 GTGAGAAATTAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200
 Oy 1304 CTCTTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1363
 Db 1201 CTCTTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
 Oy 1364 GTAGGTCTCTTACAGATTTGTGGGCGCCAGCCCGTACACAGGAGCTGTAACCTGTG 1423
 Db 1261 GTAGGTCTCTTACAGATTTGTGGGCGCCAGCCCGTACACAGGAGCTGTAACCTGTG 1320
 Oy 1424 AATCTGTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1483
 Db 1321 AATCTGTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
 Oy 1484 AAGTGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1543
 Db 1381 AAGTGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1440
 Oy 1544 ATGATTTTACAGAGTGTGTGCTACAGAGTGTGGGCTTACATCTGAGGTTGCCGATTT 1603
 Db 1441 ATGATTTTACAGAGTGTGTGCTACAGAGTGTGGGCTTACATCTGAGGTTGCCGATTT 1500
 Oy 1604 GGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1663
 Db 1501 GGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560

Oy 1664 CTACTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1723
 Db 1561 CTACTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 Oy 1724 GCGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1783
 Db 1621 TCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
 Oy 1784 AAAATCAGTATTTTACATTTCCAAAATTTGAGAGTGTGAGATTTCTTGTAGTATGAG 1843
 Db 1681 AAAATCAGTATTTTACATTTCCAAAATTTGAGAGTGTGAGATTTCTTGTAGTATGAG 1740
 Oy 1844 TTTTACGAGCTGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1903
 Db 1741 TTTTACGAGCTGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
 Oy 1904 TGTGCTTACATGAGAAATTTCAATTTGAGAAATTTGAGAAATTTGAGAAATTTGAGAA 1963
 Db 1801 TGTGCTTACATGAGAAATTTCAATTTGAGAAATTTGAGAAATTTGAGAAATTTGAGAA 1860
 Oy 1964 TTTTACGAGCTGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2023
 Db 1861 TTTTACGAGCTGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
 Oy 2024 GTTGTTTTAAATTAAGGATTCATCTCATTTACAGGTAG 2062
 Db 1921 GTTGTTTTAAATTAAGGATTCATCTCATTTACAGGTAG 1959

RESULT 5
 US-09-837-992-19
 ; Sequence 19, Application US/09837992
 ; Patent No. US20020081687A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tian, Hui
 ; APPLICANT: Schultze, Joshua
 ; APPLICANT: Shan, Bei
 ; APPLICANT: Tularek Inc.
 ; TITLE OF INVENTION: Sltosterolemia Susceptibility Gene (SSG): Compositions
 ; FILE REFERENCE: 018781-006020US
 ; CURRENT APPLICATION NUMBER: US/09/837,992
 ; PRIORITY FILING DATE: 2001-04-18
 ; PRIOR APPLICATION NUMBER: US 60/198,465
 ; PRIOR FILING DATE: 2000-04-18
 ; PRIOR APPLICATION NUMBER: US 60/204,234
 ; NUMBER OF SEQ ID NOS: 45
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 19
 ; LENGTH: 472
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: exon 13 of hSSG
 ; US-09-837-992-19

Query Match 20.2%; Score 472; DB 9; Length 472;
 Best Local Similarity 100.0%; Pred. No. 3.2e-132;
 Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1869 GCAGCTCAATGTTTCTGTGACACTAATTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1928
 Db 1 GCAGCTCAATGTTTCTGTGACACTAATTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
 Oy 1929 TCATGAGAAATACCGCCAGGAGTGCATGATGATGATGATGATGATGATGATGATGATGATG 1988
 Db 61 TCATGAGAAATACCGCCAGGAGTGCATGATGATGATGATGATGATGATGATGATGATGATG 120
 Oy 1989 ATTCATTTATTCAGAGCTTGTGATCTAGAGATGATGATGATGATGATGATGATGATGATG 2048
 Db 121 ATTCATTTATTCAGAGCTTGTGATCTAGAGATGATGATGATGATGATGATGATGATGATG 180

OY 2049 TCATTAGCAGTGAAGGAGGCTGGGAAAAATGGAAGTGAAGCTGCCAGCTGCA 2108
| | | | |
DB 181 TCATTAGCAGTGAAGGAGGCTGGGAAAAATGGAAGTGAAGCTGCCAGCTGCA 240
| | | | |
OY 2109 TGACTGCTGAGCTGGAATGAGATGAGTATTTCTTTCTTGACAGACATCTC 2168
| | | | |
DB 241 TGACTGCTGAGCTGGAATGAGATGAGTATTTCTTTCTTGACAGACATCTC 300
| | | | |
OY 2169 AAGCTTTTACATTAAGACTTCATTTGCTCTTGATGCAAGAGGCTGGAATGC 2228
| | | | |
DB 301 AAGCTTTTACATTAAGACTTCATTTGCTCTTGATGCAAGAGGCTGGAATGC 360
| | | | |
OY 2229 AATGGAAGTGTATTAAGTCCCTGCTTACAACTTCAGAGGAGATGTATTTGA 2288
| | | | |
DB 361 AATGGAAGTGTATTAAGTCCCTGCTTACAACTTCAGAGGAGATGTATTTGA 420
| | | | |
OY 2289 AATGGAAGTGTATTAAGTCCCTGCTTACAACTTCATTAATTAATTAATGAG 2340
| | | | |
DB 421 AATGGAAGTGTATTAAGTCCCTGCTTACAACTTCATTAATTAATTAATGAG 472
| | | | |

RESULT 6

US-09-837-992-7
; Sequence 7, Application US/09837992
; Patent No. US20020081687A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Hui
; APPLICANT: Schultz, Joshua
; APPLICANT: Shan, Bel
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: Sltosterolemia Susceptibility Gene (SSG): Compositions
; FILE REFERENCE: 018781-006020US
; CURRENT APPLICATION NUMBER: US/09/837,992
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/198,465
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 60/204,234
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 249
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: exon 1 of hSSG
US-09-837-992-7

Query Match 10.6%; Score 249; DB 9; Length 249;
Best Local Similarity 100.0%; Pred. No. 8.7e-65;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCAGGTGAGCAGGAGGAGCTGCGACAGGGCTCCCACTGAAGCACTTGAGGA 60
| | | | |
DB 1 GTCAGGTGAGCAGGAGGAGCTGCGACAGGGCTCCCACTGAAGCACTTGAGGA 60
| | | | |
OY 61 GGTCCGGGACACAGAAATTTGGCCAGCTTGTCTGCTGTTGGCCATGGGTACCTCTC 120
| | | | |
DB 61 GGTCCGGGACACAGAAATTTGGCCAGCTTGTCTGCTGTTGGCCATGGGTACCTCTC 120
| | | | |
OY 121 ACTTTGACCCCGGAGGCTCCATGAGTTCCTCAAGTAAACAGAGGCTCCAGAGCCCT 180
| | | | |
DB 121 ACTTTGACCCCGGAGGCTCCATGAGTTCCTCAAGTAAACAGAGGCTCCAGAGCCCT 180
| | | | |
OY 181 GAGAGGGGCTCTGCGACCGCCCGAGAGCTTCACAGCCTGGGATCTCCATCTCTCTA 240
| | | | |
DB 181 GAGAGGGGCTCTGCGACCGCCCGAGAGCTTCACAGCCTGGGATCTCCATCTCTCTA 240
| | | | |
OY 241 CAGCGTCAG 249
| | | | |
DB 241 CAGCGTCAG 249
| | | | |

RESULT 7

US-09-837-992-14
; Sequence 14, Application US/09837992
; Patent No. US20020081687A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Hui
; APPLICANT: Schultz, Joshua
; APPLICANT: Shan, Bel
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: Sltosterolemia Susceptibility Gene (SSG): Compositions
; FILE REFERENCE: 018781-006020US
; CURRENT APPLICATION NUMBER: US/09/837,992
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/198,465
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 60/204,234
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 214
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: exon 8 of hSSG
US-09-837-992-14

Query Match 9.1%; Score 214; DB 9; Length 214;
Best Local Similarity 100.0%; Pred. No. 3.2e-54;
Matches 214; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1011 TGACCTGAGCTGATGATATCCCAAGCAAGAGGAGAAATAGAAACCTCCAGAGAG 1070
| | | | |
DB 1 TGACCTGAGCTGATGATATCCCAAGCAAGAGGAGAAATAGAAACCTCCAGAGAG 60
| | | | |
OY 1071 TCCAGATGATGATATGCTGCTACAGAAATGACCAATTTGCTCAATTAATTTGAGATA 1130
| | | | |
DB 61 TCCAGATGATGATATGCTGCTACAGAAATGACCAATTTGCTCAATTAATTTGAGATA 120
| | | | |
OY 1131 TTGAAGAATGAACACCTGAAACCTTACCATGCTCTTCAAAACCAAGATTCTC 1190
| | | | |
DB 121 TTGAAGAATGAACACCTGAAACCTTACCATGCTCTTCAAAACCAAGATTCTC 180
| | | | |
OY 1191 CTGAGTTTCTCTAACTGAGTGTCTCTCTGAG 1224
| | | | |
DB 181 CTGAGTTTCTCTAACTGAGTGTCTCTCTGAG 214
| | | | |

RESULT 8

US-09-837-992-15
; Sequence 15, Application US/09837992
; Patent No. US20020081687A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Hui
; APPLICANT: Schultz, Joshua
; APPLICANT: Shan, Bel
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: Sltosterolemia Susceptibility Gene (SSG): Compositions
; FILE REFERENCE: 018781-006020US
; CURRENT APPLICATION NUMBER: US/09/837,992
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/198,465
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 60/204,234
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 206

OY 741 AGGTCATGCTGTTGATGAGCCACACAGGCGCTGAGTGCATGCTATACATG 800
DB 1 AGGTCATGCTGTTGATGATGACCAACACAGGCGCTGAGTGCATGCTATACATG 60
OY 801 TCGTCCTGCTGAGTGGAGTGCAGGACCGAATGTTGCTTCCACATTCACG 860
DB 61 TCGTCCTGCTGAGTGGAGTGCAGGACCGAATGTTGCTTCCACATTCACG 120
OY 861 CCCGTTCTGAGCTTTTCAG 880
DB 121 CCCGTTCTGAGCTTTTCAG 140

RESULT 13

US-10-027-632-152155
Sequence 152155, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Mang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIORITY FILING DATE: 2002-04-30
PRIORITY FILING DATE: 2000-07-12 US 60/218,006
PRIORITY FILING DATE: 2000-04-20 US 60/198,676
PRIORITY FILING DATE: 2000-03-29 US 60/193,483
PRIORITY FILING DATE: 2000-02-24 US 60/185,218
PRIORITY FILING DATE: 1999-11-23 US 60/167,363
PRIORITY FILING DATE: 1999-09-28 US 60/156,358
PRIORITY FILING DATE: 1999-08-09 US 60/146,002
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 152155
LENGTH: 759
TYPE: DNA
ORGANISM: Human
US-10-027-632-152155

Query Match
Best Local Similarity 99.3%; Pred. No. 2,9e-31;
Matches 139; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1431 TTCCCGTCTGCGAGCTGTTCAGCGACGAGAGTGCCTCTTACCAAGTGGC 1490
DB 42 TTCCCGTCTGCGAGCTGTTCAGCGACGAGAGTGCCTCTTACCAAGTGGC 101
OY 1491 AGATGATGCTGGCGCTTATGACGAGTGCCTCTTACCAAGTGGC 1550
DB 102 AGATGATGCTGGCGCTTATGACGAGTGCCTCTTACCAAGTGGC 161
OY 1551 TCAGCAGTGTGCTCTACTGG 1570
DB 162 TCAGCAGTGTGCTCTACTGG 181

RESULT 14

US-10-027-632-152156
Sequence 152156, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Mang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIORITY FILING DATE: 2002-04-30
PRIORITY FILING DATE: 2000-07-12 US 60/218,006

PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 152156
LENGTH: 759
TYPE: DNA
ORGANISM: Human
US-10-027-632-152156

Query Match
Best Local Similarity 99.3%; Pred. No. 2,9e-31;
Matches 139; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1431 TTCCCGTCTGCGAGCTGTTCAGCGACGAGAGTGCCTCTTACCAAGTGGC 1490
DB 42 TTCCCGTCTGCGAGCTGTTCAGCGACGAGAGTGCCTCTTACCAAGTGGC 101
OY 1491 AGATGATGCTGGCGCTTATGACGAGTGCCTCTTACCAAGTGGC 1550
DB 102 AGATGATGCTGGCGCTTATGACGAGTGCCTCTTACCAAGTGGC 161
OY 1551 TCAGCAGTGTGCTCTACTGG 1570
DB 162 TCAGCAGTGTGCTCTACTGG 181

RESULT 15

US-10-027-632-152157
Sequence 152157, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Mang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIORITY FILING DATE: 2002-04-30
PRIORITY FILING DATE: 2000-07-12 US 60/218,006
PRIORITY FILING DATE: 2000-04-20 US 60/198,676
PRIORITY FILING DATE: 2000-03-29 US 60/193,483
PRIORITY FILING DATE: 2000-02-24 US 60/185,218
PRIORITY FILING DATE: 1999-11-23 US 60/167,363
PRIORITY FILING DATE: 1999-09-28 US 60/156,358
PRIORITY FILING DATE: 1999-08-09 US 60/146,002
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 152157
LENGTH: 759
TYPE: DNA
ORGANISM: Human
US-10-027-632-152157

Query Match
Best Local Similarity 99.3%; Pred. No. 2,9e-31;
Matches 139; Conservative 1; Mismatches 0; Indels 0; Gaps 0;


```
OY      1431 TTCCGCTGCTGCAGCTGTCAGCGACCAGAGAGAGTACAGAGGCTCTACCAAGTGGC 1490
          |||
Db      42  TTCCCGTGTGCGCGAGCTGTCAGCGACCAGAGAGAGTACAGAGGCTCTACCAAGTGGC 101
          |||
OY      1491 AGATGATGCTGGCCCTATGACACTGCAGCTCCTCCCTTCAGCGTTGTTGCCACCAATGATT 1550
          |||
Db      102 AGATGATGCTGGCCCTATGACACTGCAGCTCCTCCCTTCAGCGTTGTTGCCACCAATGATT 161
          |||
OY      1551 TCAGCACTGTGTGCTACTGG 1570
          |||
Db      162 TCAGCACTGTGTGCTACTGG 181
          |||
```

Search completed: July 27, 2003, 08:56:04
Job time : 501.926 secs

PA (TULU-) TULARIK INC.
XX Tlan H, Schultz J, Shan B;
PI WPI: 2002-017598/02.
XX P-PSDB: AAL13290.
PT Novel sitosterolemia susceptibility gene polypeptide and
PT polynucleotide, useful for screening a compound that increases the
PT level of expression or activity of SSG polypeptide for treating
PT sterol-related disorder
XX
PS Claim 8; Fig 8; 105bp; English.
XX
CC The invention relates to an isolated sitosterolemia susceptibility gene
CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
CC binding cassette (ABC) family cholesterol transporter. SSG is useful
CC for identifying a compound useful in the treatment or prevention of a
CC sterol-related disorder, including sitosterolemia, hyperlipidaemia,
CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or
CC nutritional deficiencies. SSG is also useful for treating heart disease
CC associated diseases or conditions including coronary heart disease and
CC other cardiovascular diseases, and sitosterolemia-associated condition
CC including arthritis, xanthomas, and chronic haemolytic anaemia. SSG
CC expression cassette is useful in the production of transgenic non-human
CC animals. SSG genes and their homologues are useful as tools for a number
CC of applications including diagnosing sitosterolemia and other
CC cardiovascular disorders, for forensics and paternity determinations,
CC and for treating any of a large number of SSG associated diseases. The
CC present sequence is human SSG DNA. Human SSG is located on chromosome
CC 2p21.
XX
XX Sequence 2340 BP; 541 A; 601 C; 598 G; 600 T; 0 other;
SQ
Query Match 100.0%; Score 2340; DB 24; Length 2340;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GTCAGGTGAGCAGCAGGAGGAGTCTGCACGCGGCTCCCACTGAAGCCATCTGGGGA 60
DB 1 GTCAGGTGAGCAGCAGGAGGAGTCTGCACGCGGCTCCCACTGAAGCCATCTGGGGA 60
OY 61 GGGTCGGCCACCAAAAATTTGGCCAGCTTGTGCTGCTGTTGGCCATGGTGACCTCTC 120
DB 61 GGGTCGGCCACCAAAAATTTGGCCAGCTTGTGCTGCTGTTGGCCATGGTGACCTCTC 120
OY 121 ATCTTGAACCCCGGAGGCTCCATGGGTCTCCAAATTAACAGAGCTCCAGAGCTCCT 180
DB 121 ATCTTGAACCCCGGAGGCTCCATGGGTCTCCAAATTAACAGAGCTCCAGAGCTCCT 180
OY 181 GGAGGGGCTCTGTCACCGCCCGGAGGCTCAGAGCTGGGCTCATGCTCCATGCTCCTA 240
DB 181 GGAGGGGCTCTGTCACCGCCCGGAGGCTCAGAGCTGGGCTCATGCTCCATGCTCCTA 240
OY 241 CAGCCTCAGCCACCGCCTGAGGCTCTGTGGGACATCAATCTCCGCGACAGTGGAC 300
DB 241 CAGCCTCAGCCACCGCCTGAGGCTCTGTGGGACATCAATCTCCGCGACAGTGGAC 300
OY 301 CAGGAGATCTCAAAAGATGTCTCTTGTACGTGAGAGCGGCGAGATCATGTGATCT 360
DB 301 CAGGAGATCTCAAAAGATGTCTCTTGTACGTGAGAGCGGCGAGATCATGTGATCT 360
OY 361 AGGAAGCTCAGGCTCCGGAAAAACACAGCTGTGAGAGCATGTCCGGGAGGCTGGGGG 420
DB 361 AGGAAGCTCAGGCTCCGGAAAAACACAGCTGTGAGAGCATGTCCGGGAGGCTGGGGG 420
OY 421 CGCGGGGACCTTCTGCGGAGAGGTATGTGAAGCGCGGCGCTGCCCGGAGACAGTT 480
DB 421 CGCGGGGACCTTCTGCGGAGAGGTATGTGAAGCGCGGCGCTGCCCGGAGACAGTT 480
OY 481 CCAGAGCTGCTTCTCTACGTCTCTGACAGAGACACCTGCTGAGACACCTTCACGCTCG 540
DB 481 CCAGAGCTGCTTCTCTACGTCTCTGACAGAGACACCTGCTGAGACACCTTCACGCTCG 540

OY 541 CGAGAGCTGACACTACACCGCGTGTGGCATCCGCGGCAATCCGGGCTCCCTCA 600
DB 541 CGAGAGCTGACACTACACCGCGTGTGGCATCCGCGGCAATCCGGGCTCCCTCA 600
OY 601 GAAGAAGTGAAGCCGTCATGAGAGCTGAGTGAAGCCATGTGGACAGACTGAT 660
DB 601 GAAGAAGTGAAGCCGTCATGAGAGCTGAGTGAAGCCATGTGGACAGACTGAT 660
OY 661 TGGCAACTACAGTTGGGGGCAATTTCCACGGGTGAGCGCGCGGCTTCATCGCAGC 720
DB 661 TGGCAACTACAGTTGGGGGCAATTTCCACGGGTGAGCGCGCGGCTTCATCGCAGC 720
OY 721 CCAGCTGCTCCAGATCCTAAGGTCAATGCTGTTGATGAGCCACACAGGCTGACAG 780
DB 721 CCAGCTGCTCCAGATCCTAAGGTCAATGCTGTTGATGAGCCACACAGGCTGACAG 780
OY 781 CATGACTGTAATCAGATTTGTGCTCCTGCTGTAACAGGCTGCGAGACAGAAATTT 840
DB 781 CATGACTGTAATCAGATTTGTGCTCCTGCTGTAACAGGCTGCGAGACAGAAATTT 840
OY 841 GGTTCACACATTCACACAGCCCGTTCGAGCTTTTTCAGCTTTGACAAATTTGCCAT 900
DB 841 GGTTCACACATTCACACAGCCCGTTCGAGCTTTTTCAGCTTTGACAAATTTGCCAT 900
OY 901 CCGAGCTTCGAGAGGCTGATTTTCTGTGGCAGCGGAGAAATGCTGATTTCTTCA 960
DB 901 CCGAGCTTCGAGAGGCTGATTTTCTGTGGCAGCGGAGAAATGCTGATTTCTTCA 960
OY 961 TGACTGCGGTTACCTTGTGCTGAACATTCAAACCTTTGACTTATATGAGACTGAC 1020
DB 961 TGACTGCGGTTACCTTGTGCTGAACATTCAAACCTTTGACTTATATGAGACTGAC 1020
OY 1021 GTCAGTGTATCCCAAGCAGAAAGGAAATAGAAACCTTCAAGAGAGTCCAGATGAT 1080
DB 1021 GTCAGTGTATCCCAAGCAGAAAGGAAATAGAAACCTTCAAGAGAGTCCAGATGAT 1080
OY 1081 AGAATCTGCTCAAAAGAAATTCAGATTTGTATTAATTTGAAGAAATTTGAAGAA 1140
DB 1081 AGAATCTGCTCAAAAGAAATTCAGATTTGTATTAATTTGAAGAAATTTGAAGAA 1140
OY 1141 GAAACACCGAAGAAAGTATACCAATGATGCTTTCAAAAGAAATGCTCCGAGTTT 1200
DB 1141 GAAACACCGAAGAAAGTATACCAATGATGCTTTCAAAAGAAATGCTCCGAGTTT 1200
OY 1201 CTCTAATCTGGGTGTTCTCTGAGAGAGTGAAGAAATCTGTGAGAAATTAAGCTGAC 1260
DB 1201 CTCTAATCTGGGTGTTCTCTGAGAGAGTGAAGAAATCTGTGAGAAATTAAGCTGAC 1260
OY 1261 AGTGAATGCGGCTCTTCAAGATCTGATGATGATGATGATGATGATGATGATGAT 1320
DB 1261 AGTGAATGCGGCTCTTCAAGATCTGATGATGATGATGATGATGATGATGATGAT 1320
OY 1321 GCGGGTCCGAAGCAATGTGCTAAAGGCTATTCAGAGACCGGATGATGATGATGAT 1380
DB 1321 GCGGGTCCGAAGCAATGTGCTAAAGGCTATTCAGAGACCGGATGATGATGATGAT 1380
OY 1381 GTTGTGGGCGCCACCCGCTACAGAGCATGCTGAATCTGTGATGATGATGATGAT 1440
DB 1381 GTTGTGGGCGCCACCCGCTACAGAGCATGCTGAATCTGTGATGATGATGATGAT 1440
OY 1441 GCGAGCTGTCAGCGCAGAGAGAGTCAAGAGAGGCTCTACCAAGAAATGAGATGAT 1500
DB 1441 GCGAGCTGTCAGCGCAGAGAGAGTCAAGAGAGGCTCTACCAAGAAATGAGATGAT 1500
OY 1501 GGCCTATGCTACGCTGCTCCCTTACAGGCTTTTCCACACATGATTTTCCAGAGCT 1560
DB 1501 GGCCTATGCTACGCTGCTCCCTTACAGGCTTTTCCACACATGATTTTCCAGAGCT 1560
OY 1561 GTGCTACTGAGAGCTGGGCTTACATCTGAGGTTGCCGATTTGGATATTTTCTGCTGC 1620
DB 1561 GTGCTACTGAGAGCTGGGCTTACATCTGAGGTTGCCGATTTGGATATTTTCTGCTGC 1620

PD	XX	17-OCT-2002.
PR	XX	20-NOV-2001; 2001WO-US43823.
PP	XX	20-NOV-2000; 2000US-252235P.
PR	XX	28-NOV-2000; 2000US-253645P.
PA	(TULARIC INC. TEXA) UNIT TEXAS SYSTEM.	
PI	Hobbs HH, Shan B, Barnes R, Tian H;	
DR	WPI: 2003-058548/05.	
DR	P-PSDB: AAE31704.	
PT	New ABCG8 polypeptides and nucleic acids, useful for treating sterol-related disorders e.g. sitosterolemia, hypercholesterolemia, hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or nutritional deficiencies -	
XX	Claim 11, Page 77; 94pp; English.	
CC	The invention relates to ATP-binding cassette (ABC) family cholesterol transporter, ABCG8 polypeptides and polynucleotides. The invention also provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known as sitosterolemia susceptibility gene (SGL). Sequences of the invention are useful for treating or preventing sterol-related disorders such as sitosterolemia, hyperlipidaemia, hypercholesterolemia, gall stones, HDL deficiency, atherosclerosis and nutritional deficiencies. They are also useful in gene therapy. The present sequence is human ABCG5 DNA.	
SQ	Sequence 2340 BP; 541 A; 601 C; 598 G; 600 T; 0 other:	
Query Match	100.0%; Score 2340; DB 25; Length 2340;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 2340;	Conservative 0; Mismatches 0; Indels 0; Caps 0	
YY	1 GTGAGGTGGAGCAGCGAGGGCAGTCGTGCCACGGGCTCCCACAATGAAAGCCACTGTGGGGA 60	
Db	1 GTGAGGTGGAGCAGCGAGGGCAGTCGTGCCACGGGCTCCCACAATGAAAGCCACTGTGGGGA 60	
YY	61 GGGTCGGGCACCACAGAAAATTGGCCACGTTTGCTGCTGTGGCAATGGGTGACCTTC 120	
Db	61 GGGTCGGGCACCACAGAAAATTGGCCACGTTTGCTGCTGTGGCAATGGGTGACCTTC 120	
YY	121 ATCTTTGACCCCCGGAGGGTCCATGAGGTCTCCAGTAAGAAGAGGCTCCAGAGCTTCCT 180	
Db	121 ATCTTTGACCCCCGGAGGGTCCATGAGGTCTCCAGTAAGAAGAGGCTCCAGAGCTTCCT 180	
YY	181 GGAGGGGGGCTCTGGCACCGCCCGGAGGACCTCACAGCTGTGGGCATCTCCATGCTCTCTA 240	
Db	181 GGAGGGGGGCTCTGGCACCGCCCGGAGGACCTCACAGCTGTGGGCATCTCCATGCTCTCTA 240	
YY	241 CAGCGTCAGCCACCGCGTAGAGGCCCTGTGTGGAGCATCAATCTTGCAGAGAGTGGAC 300	
Db	241 CAGCGTCAGCCACCGCGTAGAGGCCCTGTGTGGAGCATCAATCTTGCAGAGAGTGGAC 300	
YY	301 CAGGCAAGATCCCAAAAGATGTCTCTTTAGCTGAGAGAGGGGCAATCATGTGCATCTCT 360	
Db	301 CAGGCAAGATCCCAAAAGATGTCTCTTTAGCTGAGAGAGGGGCAATCATGTGCATCTCT 360	
YY	361 AGGAAGCTCAGGCTCCGGGAAAAACACAGCTCTGTGAGCCATGTCGGGGAGGCTGGGGCG 420	
Db	361 AGGAAGCTCAGGCTCCGGGAAAAACACAGCTCTGTGAGCCATGTCGGGGAGGCTGGGGCG 420	
YY	421 CGCGGGGACCTTCTGTGGGGAGGTATGTAGTAAGGGCGGGCGCTGTGGCGGAGAGCTT 480	
Db	421 CGCGGGGACCTTCTGTGGGGAGGTATGTAGTAAGGGCGGGCGCTGTGGCGGAGAGCTT 480	
YY	481 CCAGGACGTCCTTCTCTACGTCCTGACAGAGGACACCTGCTGTAGCAAGCTTCACCGTGGC 540	

Db	481	CCAGAGACTGCTTCTCTCTACGTCCTCAGACGACACACCTGCTGAGCAGACCTTCACCGTGC	540
OY	541	CGAGACGCTGCACCTACACCGCGCTGCGCCATCCGCGCGGCAATCCGCGCTCTCTCCA	600
Db	541	CGAGACGCTGCACCTACACCGCGCTGCGCCATCCGCGCGGCAATCCGCGCTCTCTCCA	600
OY	601	GAAAGAGGTGGAGGCGCTCATGGCAGAGCTGAGTCTGAGCCATGTTGGCAGACCGAGT	660
Db	601	GAAAGAGGTGGAGGCGCTCATGGCAGAGCTGAGTCTGAGCCATGTTGGCAGACCGAGT	660
OY	661	TGGCAACATACAGCTTGGGGGGGCAATTCACGGGTAGAGGGCGCGCGGCTCCATGCGAC	720
Db	661	TGGCACTACTACAGCTTGGGGGGGCAATTCACGGGTAGAGGGCGCGCGGCTCCATGCGAC	720
OY	721	CCAGCTGCTCCAGATTCCTTAAGGTCATCTGTTTATGATAGCCACACAGGCTGGACGTG	780
Db	721	CCAGCTGCTCCAGATTCCTTAAGGTCATCTGTTTATGATAGCCACACAGGCTGGACGTG	780
OY	781	CATGACTGCTAATCAGATTGTGTCCTCTGGTGGAACTGGCTCGCAGAAACGAATGT	840
Db	781	CATGACTGCTAATCAGATTGTGTCCTCTGGTGGAACTGGCTCGCAGAAACGAATGT	840
OY	841	GGTTCCTACCACTTACACGAGCGCCCGCTTCGAGCTTTTACGCTTTTGACAAATTCGAT	900
Db	841	GGTTCCTACCACTTACACGAGCGCCCGCTTCGAGCTTTTACGCTTTTGACAAATTCGAT	900
OY	901	CCTGAGCTCGGAGAGCTGATTTTCTGAGACGCGACGCGGAAAGCTGATTCCTTCA	960
Db	901	CCTGAGCTCGGAGAGCTGATTTTCTGAGACGCGCGGAAAGCTGATTCCTTCA	960
OY	961	TGACTGCGGTTACCCCTTGTCTCTGAACTTCAAAACCTTTGACTCTTATGAGACCTGAC	1020
Db	961	TGACTGCGGTTACCCCTTGTCTCTGAACTTCAAAACCTTTGACTCTTATGAGACCTGAC	1020
OY	1021	GTCGTGGATTCCTCCAAAGCAGAAACGGGAAATAGAAACCTCCCAAGAGCTCCAGAT	1080
Db	1021	GTCGTGGATTCCTCCAAAGCAGAAACGGGAAATAGAAACCTCCCAAGAGCTCCAGAT	1080
OY	1081	AGAAATGCTCCACAAAGAAATCAGCAATTTGTGCATPAAACCTTGAAGAATATTGAAGAT	1140
Db	1081	AGAAATGCTCCACAAAGAAATCAGCAATTTGTGCATPAAACCTTGAAGAATATTGAAGAT	1140
OY	1141	GAAACACCTGAAAACGTTTACCATGCTTCTTCAAAACCAAGAATTCCTCTGAGATTT	1200
Db	1141	GAAACACCTGAAAACGTTTACCATGCTTCTTCAAAACCAAGAATTCCTCTGAGATTT	1200
OY	1201	CTCTAAACCTGGGCTTCTCTGAGAGAGGTGACAAAGAACTTGGTGGAGAAATAGCTGCG	1260
Db	1201	CTCTAAACCTGGGCTTCTCTGAGAGAGGTGACAAAGAACTTGGTGGAGAAATAGCTGCG	1260
OY	1261	AGTGAATACGCGTCTCTCTCAGAAATCTGATGAGTGGTTCCTCTCTCTCTCTCTCTCT	1320
Db	1261	AGTGAATACGCGTCTCTCTCAGAAATCTGATGAGTGGTTCCTCTCTCTCTCTCTCTCT	1320
OY	1321	GCGGCTCCGAAGCAATGTCTAAAGGCTCTATCCAGAGACGCGTAGGTCTCTTACCA	1380
Db	1321	GCGGCTCCGAAGCAATGTCTAAAGGCTCTATCCAGAGACGCGTAGGTCTCTTACCA	1380
OY	1381	GTTTGTGGGGGCGCCCGCTGACACAGGCATGCTAAAGCTGTGAATCTGTTCCCGTCT	1440
Db	1381	GTTTGTGGGGGCGCCCGCTGACACAGGCATGCTAAAGCTGTGAATCTGTTCCCGTCT	1440
OY	1441	GCGAGCTGTACGCGACAGAGAGTGCAGAGCGGCTCTACACAGAAAGTGGCAGATGATCT	1500
Db	1441	GCGAGCTGTACGCGACAGAGAGTGCAGAGCGGCTCTACACAGAAAGTGGCAGATGATCT	1500
OY	1501	GGGCTATGCACTGACGCTCTCTCCCTTCAACGTTGTGCCACCACTGATTTTCACAGCTG	1560
Db	1501	GGGCTATGCACTGACGCTCTCTCCCTTCAACGTTGTGCCACCACTGATTTTCACAGCTG	1560
OY	1561	GTCCTACTGAGACGCTGGGCTTACATCCTGAGAGTTGCCGGAATTTGATATTTCTGCTGC	1620
Db	1561	GTCCTACTGAGACGCTGGGCTTACATCCTGAGAGTTGCCGGAATTTGATATTTCTGCTGC	1620

QY	1621	TCCTCTGGCCCCCACCCTTAATTGGTGAATTTCTAACTTTGTGCTACTTGGTATGCTCA	1680
Db	1621	TCCTCTGGCCCCCACCCTTAATTGGTGAATTTCTAACTTTGTGCTACTTGGTATGCTCA	1680
QY	1681	AAATCCAAATATAGCAACAGGTGTGGTCTGTGTCCATATGGGGGGGTGTGTGG	1740
Db	1681	AAATCCAAATATAGCAACAGGTGTGGTCTGTGTCCATATGGGGGGGTGTGTGG	1740
QY	1741	ATCTGGATTCCTCAGAAAACATPACAAGAAATGCCCCATTCCTTTTAAATCATCAGTATTT	1800
Db	1741	ATCTGGATTCCTCAGAAAACATPACAAGAAATGCCCCATTCCTTTTAAATCATCAGTATTT	1800
QY	1801	TCATCTCCAAAATATTTGCAGTGCAGATTCCTGTGTACTCAATGAGTTCAGGACTGAAATTT	1860
Db	1801	TCATCTCCAAAATATTTGCAGTGCAGATTCCTGTGTACTCAATGAGTTCACGACTGAAATTT	1866
QY	1861	CACCTGTGGCAGCTCAAAATGTTTCTGTACACAATPACAATGAGTGTGCTTCACTCAAG	1920
Db	1861	CACCTGTGGCAGCTCAAAATGTTTCTGTACACAATPACAATGAGTGTGCTTCACTCAAG	1920
QY	1921	AATTCAAATTCATTGAGAAAACCTGCCAGGTGCAACATCTAGATTACATGAACTTTCT	1980
Db	1921	AATTCAAATTCATTGAGAAAACCTGCCAGGTGCAACATCTAGATTACATGAACTTTCT	1980
QY	1981	GATTTTGTATTCATTTATTTCCAGCTCTTGTGATCCTAGAAATGTTTGTTCAAAATPAG	2040
Db	1981	GATTTTGTATTCATTTATTTCCAGCTCTTGTGATCCTAGAAATGTTTGTTCAAAATPAG	2040
QY	2041	GGATATATCTAATTCAGAGTAGTGAAAGCCATAGCTGGGAAAATGGAAGTAACTGCCG	2100
Db	2041	GGATATATCTAATTCAGAGTAGTGAAAGCCATAGCTGGGAAAATGGAAGTAACTGCCG	2100
QY	2101	ACTGTGCAATGACTGCTCTGAAAGCTGTGAAATGAGAGTCCCATGATTTCTTTTGACAG	2160
Db	2101	ACTGTGCAATGACTGCTCTGAAAGCTGTGAAATGAGAGTCCCATGATTTCTTTTGACAG	2160
QY	2161	GACATCTCAAGTCTTTTAAACCATTAAGACTTCATTTGTGCTCTTGATCCAGAGGCC	2220
Db	2161	GACATCTCAAGTCTTTTAAACCATTAAGACTTCATTTGTGCTCTTGATCCAGAGGCC	2220
QY	2221	TTGAATGCAATGAGAAGTGGTTTATAGTCCCTGTGTACAACTTGAGGAGACATGTGGT	2280
Db	2221	TTGAATGCAATGAGAAGTGGTTTATAGTCCCTGTGTACAACTTGAGGAGACATGTGGT	2280
QY	2281	TATTTGCAAAATTTGACTGAGCGGACCCAGCAAGATGTAATTAATTCATAAACCCTATGG	2340
Db	2281	TATTTGCAAAATTTGACTGAGCGGACCCAGCAAGATGTAATTAATTCATAAACCCTATGG	2340

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATE/) PATEL S B.
 PA (DEAN/) DEAN M.
 XX Patel SB, Dean M:
 XX WPI: 2002-416483/44.
 XX
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 XX
 XX Example 3: Page 37-38; 66pp; English.
 XX
 CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present nucleic
 CC acid sequence represents the cDNA sequence of human ABCG5 gene located
 CC on chromosome 2p21.
 XX
 XX Sequence 2516 BP; 601 A; 631 C; 636 G; 648 T; 0 other;
 SQ
 Query Match 99.9%; Score 2338.4; DB 24; Length 2516;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2339; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

455 CGCGGGAGCCTTCTCGGGGAGGTATATGTAACGGCCGGCGCTGCCCGGAGCAGTT 514
 Db
 481 CCAGAGCTGCTTCTCTACAGCTCTGACAGAGCAGACACCTCTGAGACCTTCACCTGCG 540
 QY
 515 CCAGAGCTGCTTCTCTACAGCTCTGACAGAGCAGACACCTCTGAGACCTTCACCTGCG 574
 Db
 541 CGAGAGCTGCTCTACAGCTCTGACAGAGCAGACACCTCTGAGACCTTCACCTGCG 600
 QY
 575 CGAGAGCTGCTCTACAGCTCTGACAGAGCAGACACCTCTGAGACCTTCACCTGCG 634
 Db
 601 GAAGAAGCTGAGAGCCCTCATGAGAGAGCTGAGCTGAGAGCTGAGAGCTGAGAGCTGAG 660
 QY
 635 GAAGAAGCTGAGAGCCCTCATGAGAGAGCTGAGCTGAGAGCTGAGAGCTGAGAGCTGAG 694
 Db
 661 TGCCAACTACAGCTTGGGGGGCATTTTCCACGGGTGAGCCGGCGGCTTCATCCGACG 720
 QY
 695 TGCCAACTACAGCTTGGGGGGCATTTTCCACGGGTGAGCCGGCGGCTTCATCCGACG 754
 Db
 721 CCAGCTGCTCAGAGATCTTAAGTCAATGCTGCTTGTGATGACCAACACAGGCTGACCTG 780
 QY
 755 CCAGCTGCTCAGAGATCTTAAGTCAATGCTGCTTGTGATGACCAACACAGGCTGACCTG 814
 Db
 781 CATGACTGCTAATCAGATTTGTGCTCTCTGAGAACTGGCTCGGAGAACCGAATTTG 840
 QY
 815 CATGACTGCTAATCAGATTTGTGCTCTCTGAGAACTGGCTCGGAGAACCGAATTTG 874
 Db
 841 GGTTCACCATTCACAGAGCCCGTCTGAGCTTTTTCAGCTTTTTCAGCAAAATTTGCGAT 900
 QY
 875 GGTTCACCATTCACAGAGCCCGTCTGAGCTTTTTCAGCTTTTTCAGCAAAATTTGCGAT 934
 Db
 901 CCGAGCTTCCGAGAGCTGATTTTCTGTCGACAGCCAGGGAATGCTTGAATTTTTCGAA 960
 QY
 935 CCGAGCTTCCGAGAGCTGATTTTCTGTCGACAGCCAGGGAATGCTTGAATTTTTCGAA 994
 Db
 961 TGACTCGGTTTACCTTGTCTGATCAATTAACCTTTGATCTTATATGAGACCTGAC 1020
 QY
 995 TGACTCGGTTTACCTTGTCTGATCAATTAACCTTTGATCTTATATGAGACCTGAC 1054
 Db
 1021 GTCAGTGTATACCAAGCAAGAGAGGGAATAGAACTCCCAAGAGCTCAGATGAT 1080
 QY
 1055 GTCAGTGTATACCAAGCAAGAGAGGGAATAGAACTCCCAAGAGCTCAGATGAT 1114
 Db
 1081 AGAATGTGCTTACAAAGCAATTCGCAATTTTGTCTAATAACTTTGAGCAATATGAAAGAT 1140
 QY
 1115 AGAATGTGCTTACAAAGCAATTCGCAATTTTGTCTAATAACTTTGAGCAATATGAAAGAT 1174
 Db
 1141 GAACACCTGAAAGCAATTCGCAATTTTGTCTAATAACTTTGAGCAATATGAAAGAT 1200
 QY
 1175 GAACACCTGAAAGCAATTCGCAATTTTGTCTAATAACTTTGAGCAATATGAAAGAT 1234
 Db
 1201 CTCTAAACTGGGTGTTCTCTGAGAGAGTGAACAAAGCTTGTGAGAAATTAAGCTGCG 1260
 QY
 1235 CTCTAAACTGGGTGTTCTCTGAGAGAGTGAACAAAGCTTGTGAGAAATTAAGCTGCG 1294
 Db
 1261 AGTGATTAACGCTCTCTTACAGATTCATGATGATGATGATGATGATGATGATGATGAT 1320
 QY
 1295 AGTGATTAACGCTCTCTTACAGATTCATGATGATGATGATGATGATGATGATGATGAT 1354
 Db
 1321 GCGGGTCCGAAGCAATGTGCTTAAAGGCTCTATCCAGAGACCGGCTGATGCTCTTTACCA 1380
 QY
 1355 GCGGGTCCGAAGCAATGTGCTTAAAGGCTCTATCCAGAGACCGGCTGATGCTCTTTACCA 1414
 Db
 1381 GTTTGTGGGCGCACCCCTTACAGAGCTGCTGAAGAGCTGTAAGCTGTTCCCGTGCT 1440
 QY
 1415 GTTTGTGGGCGCACCCCTTACAGAGCTGCTGAAGAGCTGTAAGCTGTTCCCGTGCT 1474
 Db
 1441 GCGAGCTGTACAGCAAGCAGAGAGTACAGAGGCTTACAGAGAGTACAGAGTACAGAGTAC 1500
 QY
 1475 GCGAGCTGTACAGCAAGCAGAGAGTACAGAGGCTTACAGAGAGTACAGAGTACAGAGTAC 1534
 Db
 1501 GGCCTATGACATGACAGCTCTCCCTTACAGCTTGTGACACATGATTTTACAGAGTGT 1560
 QY
 1535 GGCCTATGACATGACAGCTCTCCCTTACAGCTTGTGACACATGATTTTACAGAGTGT 1594
 Db

QY	1561	GTGCTACTGAGACGCTGGGCTTACATCCTGAGGTTGCCGATTTGGATATTTTTCGCTGC	1620
Db	1595	GTGCTACTGAGACGCTGGGCTTACATCCTGAGGTTGCCGATTTGGATATTTTTCGCTGC	1654
QY	1621	TCCTGTGGCCCCACCTTAATTTGGGAAATTTCTAACTCTGTGCTACTGTGTAATCTCCA	1680
Db	1655	TCCTGTGGCCCCACCTTAATTTGGGAAATTTCTAACTCTGTGCTACTGTGTAATCTCCA	1714
QY	1681	AAATCCAAATTAATGATCAACAGTGTAGTGGCTCTGCTGCATTCGCGGGGGGCTGTGTTGG	1740
Db	1715	AAATCCAAATTAATGATCAACAGTGTAGTGGCTCTGCTGCATTCGCGGGGGGCTGTGTTGG	1774
QY	1741	ATCTGGATTCCTCGAAGAACATACAGAAATGCCATTCCTTTTAAATTCATCACTTAATT	1800
Db	1775	ATCTGGATTCCTCGAAGAACATACAGAAATGCCATTCCTTTTAAATTCATCACTTAATT	1834
QY	1801	TACATTTCCAAAATATTTGGCAGTGGGATTCCTGTATGTCAATGAGATTTTACGAGCTGAATTT	1860
Db	1835	TACATTTCCAAAATATTTGGCAGTGGGATTCCTGTATGTCAATGAGATTTTACGAGCTGAATTT	1894
QY	1861	CACATTGTGGACGCTCAATATGTTTCTGTGACAACTAACTCAATGTGCTCCTTCACTCAAGG	1920
Db	1895	CACATTGTGGACGCTCAATATGTTTCTGTGACAACTAACTCAATGTGCTCCTTCACTCAAGG	1954
QY	1921	AATTCATTTCACTTGGAGAAAACCTGCCAGGTGCACACTCTAGATTTCACAAATGAACCTTCT	1980
Db	1955	AATTCATTTCACTTGGAGAAAACCTGCCAGGTGCACACTCTAGATTTCACAAATGAACCTTCT	2014
QY	1981	GATTTTGTATTTCAATTTATTCACAGCTCTGTGCATCCTTAGGAAATAGTGTTTTTCAAATTAAG	2040
Db	2015	GATTTTGTATTTCAATTTATTCACAGCTCTGTGCATCCTTAGGAAATAGTGTTTTTCAAATTAAG	2074
QY	2041	GGATTCATCTCATTTAGCAGTAGTGAAGAACCATGTGGCTGGGAAAAATGGAAGTGAAGCTCCG	2100
Db	2075	GGATTCATCTCATTTAGCAGTAGTGAAGAACCATGTGGCTGGGAAAAATGGAAGTGAAGCTCCG	2134
QY	2101	ACTGTGCATACACGCTCTGAAACGCTCTGAAATGAGAGTGCATGTATTTCTTCTTGACAG	2160
Db	2135	ACTGTGCATACACGCTCTGAAACGCTCTGAAATGAGAGTGCATGTATTTCTTCTTGACAG	2194
QY	2161	GACATCTCAAGTCTTTTAAACATTAAGATCAATTTTGTGCTCTGTTGGATCCAGAGGCG	2220
Db	2195	GACATCTCAAGTCTTTTAAACATTAAGATCAATTTTGTGCTCTGTTGGATCCAGAGGCG	2254
QY	2221	TTGAAATGCAATGGAAGTGGTTTATGTATTCCTTCTCTTACACATTGCGAGGAGATGTGT	2280
Db	2255	TTGAAATGCAATGGAAGTGGTTTATGTATTCCTTCTCTTACACATTGCGAGGAGATGTGT	2314
QY	2281	TATTTGGAAATTTGAGCTAGCGGCGGACCCAGAAATGTATTAATATTTCTATTAACCTATAGG	2340
Db	2315	TATTTGGAAATTTGAGCTAGCGGCGGACCCAGAAATGTATTAATATTTCTATTAACCTATAGG	2374

	RESULT 4
	ABK51681
ID	ABK51681 standard; DNA; 1920 BP.
XX	
AC	ABK51681;
XX	
DT	30-JUL-2002 (first entry)
XX	
DE	DNA encoding human ABCG5 protein.
XX	
KW	Human: ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
KW	arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
KW	chromosome 2p21; ds.
XX	
OS	Homo sapiens.
XX	
FT	Key
FT	Location/Qualifiers
CDS	1..1920
FT	/tag- a

FT	/product="Human ABCG5 protein"
FT	/transl_except- (pos: 4..9, aa: GDLSITPQSGMGL)
FT	/note- "This sequence contains 13 exons"
XX	
XX	
XX	
PD	04-APR-2002.
XX	
XX	
XX	
PR	25-SEP-2001; 2001WO-US29859.
XX	
XX	25-SEP-2000; 2000US-235268P.
PA	(US\$) US DEPT HEALTH & HUMAN SERVICES.
PA	(PATEL/) PATEL S B.
PA	(DEAN/) DEAN M.
XX	
PI	Patel SB, Dean M;
XX	
XX	WPI: 2002-416483/44.
DR	P-PSDB: AAU98984.
PT	
PT	Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
PT	nucleic acid encoding the polypeptide, useful for treating
PT	stereosterolemia, arteriosclerosis and heart diseases
XX	
XX	
PS	Claim 38; Page 36-37; 66pp; English.
XX	
CC	The present invention relates to a new mammalian ATP-binding cassette
CC	gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC	predisposition for developing stereosterolemia, arteriosclerosis or heart
CC	disease. The molecules of the invention are also useful for identifying
CC	a compound which alters ABCG5 activity level comprising contacting a cell
CC	culture or mammal which have ABCG5 polypeptide with a compound and
CC	measuring ABCG5 biological activity in the cell culture or in mammal,
CC	where an increase or decrease in ABCG5 biological activity compared to
CC	ABCG5 biological activity in a control cell culture or mammal not
CC	contacted with the compound, identifies a compound that increases or
CC	decreases ABCG5 activity respectively. The cell culture or mammal
CC	comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC	ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC	polypeptide in a cell culture or mammal is also compared with that of a
CC	second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC	Stimulation of ABCG5 activity is useful for treating or preventing
CC	hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC	disease. The method of the invention is useful for increasing cholesterol
CC	excretion and/or decreasing cholesterol adsorption. The present nucleic
CC	acid sequence represents the human ABCG5 gene located on chromosome 2p21.
CC	This sequence encodes the human ABCG5 protein of the invention.
XX	
XX	
SQ	Sequence 1920 BP; 440 A; 503 C; 486 G; 491 T; 0 other;
Query Match	82.1%; Score 1920; DB 24; Length 1920;
Best Local Similarity	100.0%; Prod. No. 0;
Matches 1920;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
143	ATGGGCTCCAGTAACAGAGGCTCCACAGAGTCCCTGTGAGGGGGCTCTGCCACCGCC 202
1	ATGGGTCTCCCAAGTAACAGAGGCTCCACAGAGTCCCTGTGAGGGGGCTCTGCCACCGCC 60
203	CCGAGGCTCAACAGCCCTGGGCGATCTCCATGCTCTCAACAGCTCAGCCACCGCGTAGG 262
61	CCGAGGCTCAACAGCCCTGGGCGATCTCCATGCTCTCAACAGCTCAGCCACCGCGTAGG 120
263	CCCTGTGAGGAGCATCATCATCTTGGCCGGCAGACAGTGGAGCCAGGAGATCTCAAGATGTC 322
121	CCCTGTGAGGAGCATCATCATCTTGGCCGGCAGACAGTGGAGCCAGGAGATCTCAAGATGTC 180
323	TCTCTTGTAGCTGTGAGAGCGGCGAGATCATGTGATCTCTAGGAAGCTCAGGCTCCGGGAAA 382
181	TCTCTTGTAGCTGTGAGAGCGGCGAGATCATGTGATCTCTAGGAAGCTCAGGCTCCGGGAAA 240
383	ACACAGCTGTGTGAGAGCGGCGAGATCATGTGAGAGCTGGGGCGCGGGGACCTTCTCTGGGGAG 442

Db 241 ACCAGCTGCTGAGAGCCATGTCCGGAGAGCTGGGGGCGGGGACCTTCTGGGGAG 300
 Oy 443 GTGATGATGAAAGGCGCGCGCTGCGCGGAGAGATTCAGAGCTGCTTCTAGATC 502
 Db 301 GTGATGATGAAAGGCGCGCGCTGCGCGGAGAGATTCAGAGCTGCTTCTAGATC 360
 Oy 503 CTGAGAGAGCAGACCTGCTGAGAGAGCTCAGCGGCGGAGAGCTGCTTCTAGATC 562
 Db 361 CTGAGAGAGCAGACCTGCTGAGAGAGCTCAGCGGCGGAGAGCTGCTTCTAGATC 420
 Oy 563 CTGCTGAGAGCTGCGCGCGGAGAGCTGCTTCTAGAGAGAGAGAGAGAGAGAGAG 622
 Db 421 CTGCTGAGAGCTGCGCGGCGGAGAGCTGCTTCTAGAGAGAGAGAGAGAGAGAG 480
 Oy 623 GCAGAGCTGATGAGAGAGCTGCTGAGAGAGCTGCTTCTAGAGAGAGAGAGAGAG 682
 Db 481 GCAGAGCTGATGAGAGAGCTGCTGAGAGAGCTGCTTCTAGAGAGAGAGAGAGAG 540
 Oy 683 ATTTCAG 742
 Db 541 ATTTCAG 600
 Oy 743 GTGATGCTGCTGAG 802
 Db 601 GTGATGCTGCTGAG 660
 Oy 803 GTGCTGCTGCTGAG 862
 Db 661 GTGCTGCTGCTGAG 720
 Oy 863 GTGCTGAG 922
 Db 721 GTGCTGAG 780
 Oy 923 TTTCTGAG 982
 Db 781 TTTCTGAG 840
 Oy 983 GAAATTCAG 1042
 Db 841 GAAATTCAG 900
 Oy 1043 GAAAGGAG 1102
 Db 901 GAAAGGAG 960
 Oy 1103 GAAATTCAG 1162
 Db 961 GAAATTCAG 1020
 Oy 1163 ATGCTGCTGCTGAG 1222
 Db 1021 ATGCTGCTGCTGAG 1080
 Oy 1223 AGGAG 1282
 Db 1081 AGGAG 1140
 Oy 1283 AATGATGATGAG 1342
 Db 1141 AATGATGATGAG 1200
 Oy 1343 AAGGAG 1402
 Db 1201 AAGGAG 1260
 Oy 1403 AAGGAG 1462
 Db 1261 AAGGAG 1320
 Oy 1463 ACTGAG 1522
 Db 1321 ACTGAG 1380

Oy 1523 CCCTTCAG 1582
 Db 1381 CCCTTCAG 1440
 Oy 1583 CAGCTGAG 1642
 Db 1441 CAGCTGAG 1500
 Oy 1643 GAGTATGATGAG 1702
 Db 1501 GAGTATGATGAG 1560
 Oy 1703 GAGTATGATGAG 1762
 Db 1561 GAGTATGATGAG 1620
 Oy 1763 CAGGAG 1822
 Db 1621 CAGGAG 1680
 Oy 1823 GAGTATGATGAG 1882
 Db 1681 GAGTATGATGAG 1740
 Oy 1883 TCTGAG 1942
 Db 1741 TCTGAG 1800
 Oy 1943 TGCCAG 2002
 Db 1801 TGCCAG 1860
 Oy 2003 GCTCTGATGAG 2062
 Db 1861 GCTCTGATGAG 1920

RESULT 5
 ID ABK51685 standard; cDNA; 2354 BP.
 AC ABK51685;
 XX 30-JUL-2002 (first entry)
 DT Mouse ABCG5 cDNA sequence.
 DE
 XX Mouse; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
 KM arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
 KW ss.
 XX Mus sp.
 Os
 W0200227016-A2.
 PD 04-APR-2002.
 XX 25-SEP-2001; 2001WO-029859.
 PP 25-SEP-2000; 2000US-235268P.
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (PATEL/) PATEL S B.
 PA (DEAN/) DEAN M.
 XX Patel SB, Dean M.
 PI WPI: 2002-416483/44.
 DR Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases

XX Example 3; Page 45; 66pp; English.

PS The present invention relates to a new mammalian ATP-binding cassette
XX gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, arteriosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying
CC a compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a
CC stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present nucleic
CC acid sequence represents the cDNA sequence of the mouse ABCG5 gene of
XX the invention.

XX Sequence 2354 BP; 573 A; 604 C; 594 G; 583 T; 0 other;

Query Match 60.2%; Score 1409.2; DB 24; Length 2354;
Best Local Similarity 80.4%; Pred. No. 0;
Matches 1664; Conservative 0; Mismatches 403; Indels 3; Gaps 1;

XX 25 CTGCCAGGCGCTCCCAACTGAAAGCACTCTGGGAGGTCGCGGACCCAGAAATTTGC 84
XX 57 CTCCCAATGGCTCTCTAGTTAAAGTCGCCCTGGAGCGGACAGGCCATGAAATTTAC 116
XX 85 CCAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 144
XX 117 TTCCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 176
XX 145 GGGTCTCAAGTAAAGAGAGGCTCCAGAGCTCCCTGGAGGGGCTCTGCGACCGCC 204
XX 177 AAGGCGCTCACTCAAGAGAGAGGCTCTGAGCTCCCTGGAGGAGTTCCGTCACGGGAC 236
XX 205 GAGAGCT--CACAGCTGGGATCTCTCATGCTCTCTACAGGCTGAGCCAGCCGTGAG 261
XX 237 AGAGGCTCGGACAGCTAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 296
XX 262 GCGCTGTGGAGATACATCTTCCGCGAGAGCTGAGCCAGGACAGTCTCAAGATGT 321
XX 297 GCGTGTGGAGATCAATATCATGACAGCAAGAGTGGAGACAGCAATCTCTCAAGATGT 356
XX 322 CTCTGTAGTGGAGAGGCGGAGATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 381
XX 357 CTCTGTAGTGGAGAGGCGGAGATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 416
XX 382 AACCAAGCTGTGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 441
XX 417 GACCAAGCTGTGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 476
XX 442 GGTGTGTGGAGAGGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 501
XX 477 GGTGTGTGGAGAGGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 536
XX 502 CTGCAAGAGGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 561
XX 537 CTGCAAGAGGAGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 596
XX 562 GCT 621
XX 597 GATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 656
XX 622 GCGAGAGCTGAGTGTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 681

Db 657 GACAGAGCTGAGCTGAGCCAGCTGCGGAGCAAAATGATTTGCACTAATATTTGGGGG 716
OY 682 CATTCAGGAGGAGGAGGCGCGGCTCTCATTCGAGCCAGCTGCTCAGATCTTA 741
Db 717 AATTTCCAGTGGAGGCGCGGAGCTCTCATTCGAGCCAGCTGCTCAGATCTTA 776
OY 742 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 801
Db 777 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 836
OY 802 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 861
Db 837 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 896
OY 862 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 921
Db 897 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 956
OY 922 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 981
Db 957 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1016
OY 982 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1041
Db 1017 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1076
OY 1042 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1101
Db 1077 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1136
OY 1102 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1161
Db 1137 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1196
OY 1162 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1221
Db 1197 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1256
OY 1222 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1281
Db 1257 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1316
OY 1282 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1341
Db 1317 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1376
OY 1342 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1401
Db 1377 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1436
OY 1402 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1461
Db 1437 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1496
OY 1462 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1521
Db 1497 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1556
OY 1522 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1581
Db 1557 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1616
OY 1582 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1641
Db 1617 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1676
OY 1642 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1701
Db 1677 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1736
OY 1702 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1761
Db 1737 GGTGATGCTGTTGATGAGCAACAGAGCTGAGTGCATGATGATGATGATGATGATGAT 1796

RESULT 6
AAD22008
ID AAD22008 standard; DNA; 2258 BP.
yy

Mouse sitosterolaemia susceptibility gene (SSG).

KM mouse; atherosclerosis susceptibility gene; SSG; atherosclerosis;
KM sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; therapy;
KM gall stone; coronary heart disease; cardiovascular disease; arthritis;
KM xanthoma; haemolytic anaemia; transgenic animal; chromosome 17; ds.

05 Mus sp.

Key	Location/Qualifiers
FH	47..2005
FT	/*tag= a
CDS	/product= "Mouse Ssg protein"
FT	
FT	

WO200179272-A2.

PD 25-OCT-2001.

PF 18-APR-2001; 2001WO-US12758.

PR 18-APR-2000; 2000US-198465P.
PR 15-MAY-2000; 2000US-204234P.

PA (TULLA-) . TULARIK INC.

PI Tian H, Schultz J, Shan B;
YY

DR WPI; 2002-017598/02.

XX
XX
Normal

PT Novel *stt* sterolemia susceptibility gene polypeptide and polynucleotide, useful for screening a compound that increases the PT level of expression or activity of Sso polypeptide for treating sterol-related disorder

PS Claim 8; Fig 7; 105pp; English.

CC The invention relates to an isolated Sitosterolemia Susceptibility Gene
CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)

CC bindingcassette (ABC) family cholesterol transporter. SSG is useful
CC for identifying a compound useful in the treatment or prevention of a
CC sterol-related disorder, including sitosterolaemia, hypellipidaemia,
CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or
CC nutritional deficiencies. SSG is also useful for treating cholesterol-
CC associated diseases or conditions including coronary heart disease and
CC other cardiovascular diseases, and sitosterolaemia-associated condition
CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
CC expression cassette is useful in the production of transgenic non-human
CC animals. SSG genes and their homologues are useful as tools for a number
CC of applications including diagnosing sitosterolaemia and other
CC cardiovascular disorders, for forensics and paternity determinations,
CC and for treating any of a large number of SSG associated diseases. The
CC present sequence is mouse SSG DNA. Mouse SSG is located on chromosome 17.
XX
50 Sequence 2258 BP; 549 A; 579 C; 567 G; 563 T; 0 other:

Query Match	59.6%	Score 1395.6	DB 24	Length 2258
Best Local Similarity	80.7%	Pred. No. 0		
Matches 1642; Conservative	0	Mismatches 389	Indels 3	Gaps 1

[illegible]

121ATGCTTGACCCCGAGAGGCTCATTGGGTCTCCAGTAACACAGAGGCTCCAGAGCTCCCT 180
 61 CTTCTGAGTCCAGAGGAGGCCAGAGGSCCTCACATCAACAGAGGGTCTCTGAGCTCCCT 120
 Db

Db 121 GGAGCAAGTTCGTCACGGCAGAGGCTCGGCAAGCTTAGGTGTCTCGCATGTGTC 180

250 CTACAGCGTCAGCCACCAGCGTGAGGCCCTGGTGGACATCACATCTTCCCGGACGACGTG 297

[illegible]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

CC BY-NC-ND 4.0 International license.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

019 CCCCCCCCCCCCMCCCCC-----

[illegible]

0v 478

[illegible][illegible]

D_b

481 GCGCAGACGTTCGATACACAAGCCGATGCTGGCCCTTGCCTGCCCACTTTCTTA 540

598 CCAGAGAGGTGGAGGCCGTCATGGCAGAGCTGAGTCTGAGGCCATGTGGCAGACCGACT 657

Db 541 CAACAGAGGTAGAGGCAGTCATGACAGAGCTGAGCCTGAGCCACCGTGGCGGACCAAT 600

658 GATTGGCACTACAGCTTGGGGGCATTTCACGCGGTGAGCGCGCGCGGTCTCCATCGC 717

Db 601 GATTGCAGCTATATTTTGGGGAATTTCCAGTGGCGAGCGCGCCGAGTTCCATCGC 660

718 AGCCAGCTGCTCCAGATCCTAAGTCAATGCTGTTGATGAGCCAACACAGGCTTGA 777

Db 661 AGCCCACTCCTCAGGACCCCAAGTGCATGCTAGATGAGCCCAACACAGGACTGGA 720

778 CTGCATGACTGCTAATCAGATTGTGCTCTCTGCTGGAAGTGGCTGCAGAACCGAAT 837

Db 721 CTGCATGACTGCAAATCAAAATTGTCCTTCTCTTGCGTGAGCTGGCTCGCAGGACCGAAT 780

```

OY 838 TGTGGTCTCACATTTCAACACCCCGTTCGAGCTTTTTCAGCTTTTGACAAAATTCG 897
    ||||| || ||||| ||||| || ||||| || ||||| || ||||| |||||
DB 781 TGTGATGTTCACATTCACACACCGCTCGCTGAGCTTTCACAACTTCGACAAAATTCG 840
OY 898 CATCTGAGCTTCGAGAGCTGATTTTCTGTGACGACGAGGAGAAATCTGATTTCT 957
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| |||||
DB 841 CATCTGAGCTTCGAGAGCTGATTTTCTGTGACGACGAGGAGAAATCTGATTTCT 900
OY 958 CAAATGACGAGGTTACCCCTTCTGTAACATTTCAAAACCTTTGACTTCTATATGAGCT 1017
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| |||||
DB 901 CAAATGACGAGGTTACCCCTTCTGTAACATTTCAAAACCTTTGACTTCTATATGAGCT 960
OY 1018 GACGTCAGTGTATACCCAAAGCAAGAAAGGAAATAGAAACCTTCAGAGAGTCCAGAT 1077
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| |||||
DB 961 GACATCAGTGTATACCCAAAGCAAGAAAGGAGGAGAAATAGTACAAAGCAGTACAGAT 1020
OY 1078 GATGAGATGCTGCTTCAAGAAATGACAAATTTGTATATAAATTTGAAATATGAAAG 1137
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| |||||
DB 1021 GCTGGAATGCTCTTCAAGAAATGTAACATTTATACAAATTTCTGGAATCATGAAAG 1080
OY 1138 AATGAAACACCTGAAAGCTTACCAATGATGCTTCAAAACCAAGAAATTTCTGAGAT 1197
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| |||||
DB 1081 ACGAGATATACCTGAAAGCTTACCAATGATGCTTCAAAACCAAGAAATTTCTGAGAT 1140
OY 1198 TTTTCTTAACCTGGGCTTCTCTGAGAGAGTGAACAAGAACTTGTGAGAAATAGCT 1257
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| |||||
DB 1141 GTTCGCAAGCTTGGTGTCTGCTGAGGAGGATACAAAGAACTTAATGAGAGATTAAGCA 1200
OY 1258 GGCAGTATGAGGCTGCTCCTCAAAATCTGATGATGAGGTTGTCTCTCTTTTCTTCT 1317
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| |||||
DB 1201 GGCAGTATGATGCTGCTCCTCAAAATCTGATGATGAGGCTTCTCTCTTTTCTTCT 1260
OY 1318 TCTGCGGGTCCGAAACCAATGATGCTTAAAGGATGCTTATCCAGAGAGGCTGAGCTCTTA 1377
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| |||||
DB 1261 TCTGCGGGTCCGAAACCAATGATGCTTAAAGGATGCTTATCCAGAGAGGCTGAGCTCTTA 1320
OY 1378 CCAATTTTGGGCGCCACCCGCTACAGAGCAATGCTGAACGCTGTAATCTGTTTCCCT 1437
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| |||||
DB 1321 TCAGCTTGTGGGTGCGACCCCATACAGCGCATGCTCAATGCTGTAATCTGTTTCCAT 1380
OY 1438 GCTGAGGCTGTCACAGCAACAGAGAGAGTCAAGAGGCTCTTACCAAGAAATGGCAGATAT 1497
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| |||||
DB 1381 GCTGAGAGCCGTCACAGCAACAGAGAGAGTCAAGAGGCTCTTACCAAGAAATGGCAGATAT 1440
OY 1498 GCTGAGGCTGTCACAGCAACAGAGAGAGTCAAGAGGCTCTTACCAAGAAATGGCAGATAT 1557
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| |||||
DB 1441 GCTGAGGCTGTCACAGCAACAGAGAGAGTCAAGAGGCTCTTACCAAGAAATGGCAGATAT 1500
OY 1558 TGTGCTGATGAGAGCTGGGCTTACATCTGAGGTTGCGCCGATTTGGATATTTTCTGC 1617
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| |||||
DB 1501 TGTGCTGATGAGAGCTGGGCTTACATCTGAGGTTGCGCCGATTTGGATATTTTCTGC 1560
OY 1618 TGTGCTGATGAGAGCTGGGCTTACATCTGAGGTTGCGCCGATTTGGATATTTTCTGC 1677
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| |||||
DB 1561 TGTGCTGATGAGAGCTGGGCTTACATCTGAGGTTGCGCCGATTTGGATATTTTCTGC 1620
OY 1678 CCAAAATCCAAATATAGTCAACAGAGTGTGCTGTCATGCTGCGGGGCTGCTGT 1737
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| |||||
DB 1621 CCAAAATCCAAATATAGTCAACAGAGTGTGCTGTCATGCTGCGGGGCTGCTGT 1680
OY 1738 TGTGCTGATGAGAGCTGGGCTTACATCTGAGGTTGCGCCGATTTGGATATTTTCTGC 1797
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| |||||
DB 1681 TGTGCTGATGAGAGCTGGGCTTACATCTGAGGTTGCGCCGATTTGGATATTTTCTGC 1740
OY 1798 TTTTCAATTTCCAAAATATGTCAGAGTATCTTGTAGTCAATAGTCTACAGAGTAA 1857
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| |||||
DB 1741 TTTTCAATTTCCAAAATATGTCAGAGTATCTTGTAGTCAATAGTCTACAGAGTAA 1800
OY 1858 TTTTCAATTTCCAAAATATGTCAGAGTATCTTGTAGTCAATAGTCTACAGAGTAA 1917
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| |||||
DB 1801 CTTGACCTTGTGTGATGATCAACACCTCTATCTTAATACACCGATGTGCGCATCACCCA 1860
OY 1918 AGGAATTCATTCATGAGAAACCTGCCAGGTGCAACATCTAGATTCACAAATGAATCT 1977

```

```

DB 1861 AGGGGTCCAGTTCATGAGAGAAACCGCCAGAGTGTATCATTCAGATTCACGCAAACTT 1920
OY 1978 TCTGATTTGTATTCATTTATTCAGCTTCTGATCTCAGAAATAGTGTTCAAAT 2037
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| |||||
DB 1921 CCTCATCTTATATGAGGTTATATCCAGCTCTGCTATCCAGAAATAGTATTTTAAAT 1980
OY 2038 AAGGATCATCTCATTCAGAGTATGAGAAAGCATGAGGAGGAAATGAAATG 2091
    ||||| || ||||| || ||||| || ||||| || ||||| || ||||| |||||
DB 1981 CAGGACATCTGATTCAGATATAGTAAAGATGACAGCGAGAAAGGTTAATG 2034

```

```

RESULT 7
AAD48880
ID AAD48880 standard; DNA; 1959 BP.
XX
AC AAD48880;
XX
DT 24-MAR-2003 (first entry)
XX
DE Mouse ABCG5 DNA.
XX
KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;
KW sitosterolemia; hyperlipidaemia; hypercholesterolemia; gall stone;
KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;
KW mouse; ATP-binding cassette; sitosterolemia susceptibility gene; SSG;
KW ABCG5; gene; ds.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 1..1591
FT /tag="a
FT /product="mABCG5 protein"
XX
PN WO200281691-A2.
XX
PD 17-OCT-2002.
XX
PF 20-NOV-2001; 2001WO-US43823.
XX
PR 20-NOV-2000; 2000US-252235P.
XX
PR 28-NOV-2000; 2000US-253645P.
XX
PA (TULU-) TULARK INC.
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Hobbs HH, Shan B, Barnes R, Tian H;
PI WPI: 2003-058548/05.
PI P-PSDB: AAB31702.
XX
DR New ABCG8 polypeptides and nucleic acids, useful for treating
DR sterol-related disorders e.g. sitosterolemia, hypercholesterolemia,
DR hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or
DR nutritional deficiencies -
XX
PS Claim 11; Page 73; 94pp; English.
XX
CC The invention relates to ATP-binding cassette (ABC) family cholesterol
CC transporter, ABCG8 polypeptides and polynucleotides. The invention also
CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also
CC known as sitosterolemia susceptibility gene (SSG). Sequences of the
CC invention are useful for treating or preventing sterol-related disorders
CC such as sitosterolemia, hyperlipidaemia, hypercholesterolemia, gall
CC stones, HDL deficiency, atherosclerosis and nutritional deficiencies.
CC They are also useful in gene therapy. The present sequence is mouse
CC ABCG5 DNA.
XX
SQ Sequence 1959 BP; 468 A; 506 C; 495 G; 490 T; 0 other;

```

Query Match 58.4%; Score 1365.4; DB 25; Length 1959;
Best Local Similarity 81.4%; Pred. No. 0;

Query Match	Best Local Similarity	Score	DB	Length
Matches 1560; Conservative 0; Mismatches 352; Indels 3; Gaps 1	81.35;	1335.8;	24;	1915;
107 ATGGGATCCTTCATCTTGACCCCGGAGGATCCATGGGTCCCAAGTAACGAGGC	166			
1 ATGGGTAGCTGCCCTTCTTGATGTCAGAGGAGGACCAAGGCGCTCACATCAACGAGGG	60			
167 TCCGAGAGCTCCCTGAGGGGGCTCTGACCACCCCGGAGCCT--CACAGCCTTGCG	223			
61 TCTCTGAGCTCCCTGAGGAAAGTTCCGGTACGGGACAGAGAGGCTCGGACAGCTTAAGT	120			
224 ATCTCATGTCCTTCATGACGCTTCAGCCACCGCTGAGAGCCCTGGTGGGACATCATCT	283			
121 GTCTGTGATGTGTCTACACGCTGACGAACCGTTCGGGCTGGTGGACATCAATCA	180			
284 TCCGAGCAGCAGTGAACGAGCAGATCTCAAAATGATCTCTCTGACGTGAGAGCGGG	343			
181 TCCGAGCAGAGAGTGGGACAGCAAGCAATCTCAAAATGTCTCTCTGACGTGAGAGTGGC	240			
344 CAGATATGTGATCTCTAGGAGCTCAGGCTCGGGAAACACGCTGCTGAGCGCCATG	403			
241 CAGATATGTGATCTCTAGGAGCTCAGGCTCAGGAGAAACACGCTGCTGAGCGCCATG	300			
404 TCCGAGCAGCAGTGGGAGCGGCGGGAGCTCTCTGAGGAGGATGTAATCAAGCGCGGGCG	463			


```

Db      301  TCCGGAGGCTGCGGGCCACTGGGACCTCTGGAGGGAGGTGTTTGTAATGGCTGGAG 360
OY      464  CTGGCGCGGGAGCAGATTCCAGAGACTGCTTCTTCTACCTCTGACAGAGCAGACCTGCTG 523
Db      361  CTGGCGAGGAGCAGATTCCAGAGACTGCTTCTTCTACCTCTGACAGAGCAGACCTGCTG 420
OY      524  AGCAGCTTCACCGTGGCGGAGAGCGCTCCTACACCGCGCTGCTGCGCATCCGCGCGG 583
Db      421  AGCAGCTTCACCGTGGCGGAGAGCGTTCGATACACAGATGCTGGCGCTCTGCGCGAGC 480
OY      584  AATCCGCGCTCTTCCAGAAAGGTGAGGCGGTCTATGGCAGACGATGATCGAGCAT 643
Db      481  TCCGCGGACTTCTTCAACAAGAAAGGTAGAGGCACTGATGACAGAGCTGAGCTGAGCAC 540
OY      644  GTGGCAGACCGACTGATTGGCAACTACAGCTTGGGGGGCATTTCACAGGGTGGCGGCG 703
Db      541  GTGGCGGACCAAAATGATGGCAGCTAATTTTGGGGGAATTTCCAGTGGCGAGCGGCG 600
OY      704  CGGGTCTTCATCCGAGCGCCAGCTGCTCCAGATCTTAAGTCAATGCTGTTTGAAGCCA 763
Db      601  CGAGTTCATCCGAGCGCCAACTCTTCAGAGCCCAAGGTCAATGATGATGAGCCA 660
OY      764  ACCAGAGCGCTGGAGCTGATGAGCTTAATCAGATTTGCTCTCTCCGTTGGAGAACTGCT 823
Db      661  ACCAGAGCGACTGGAGCTGATGAGCTCAAAATCAAAATGCTCTCTGGCTGAGCTGCT 720
OY      824  CGCAGAAACCGAAATGCTGCTTCCACCATTCACAGCGCCGCTTCTGAGCTTTTTCAGCT 883
Db      721  CGCAGGAGCGGAATGATGATGCTTCCACCATTCACAGCGCTGCTGAGCTTTTTCAGAC 780
OY      884  TTTGACAAATTTCCATCTCGAGCTTCGAGAGCTGATTTTCTGTGGCAGCGCAGGAA 943
Db      781  TTTGACAAATTTCCATCTCGAGCTTCGAGAGCTGATTTTCTGTGGCAGCGCAGGAG 840
OY      944  ATGCTGATTTTCTCAATGATGATGCTGCGGTAACTCTTCTCTGTAATTCAAACCTTTTAC 1003
Db      841  ATGCTGCTGCTTCTCAATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
OY      1004  TTTCAATGAGACCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1063
Db      901  TTTTACATGAGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
OY      1064  AAGAGAGCTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1123
Db      961  AAGAGAGCTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
OY      1124  AAGATATTGAAGAAATGAAGACCTGGAAGAGCTTCAATGATGATGATGATGATGATGAT 1183
Db      1021  GAGAACATTGAAGAAATGAAGACCTGGAAGAGCTTCAATGATGATGATGATGATGATGAT 1080
OY      1184  GATTCCTGAGAGCTTCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1243
Db      1081  GATTCCTGAGAGCTTCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
OY      1244  GTRAGAAATTAAGCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1303
Db      1141  ATGAGAAATTAAGCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
OY      1304  CTCCTTTTCTGCTTGGCGGGTCCGAAGCAATGCTTAAGGCTCTAATCCAGAGCCG 1363
Db      1201  CTCATTTTCTACCTTTCGCGCTCCAGAACAAACCTTAAGGCGCTGTGCGAGGAGCCG 1260
OY      1364  GTAAGTCTCTTACCAAGTTTGTGGCGGACCCCTTACACAGAGCTGCTGAACGCTGG 1423
Db      1261  GTGGGCTGCTCTATCAGCTGTGGGTGACCCCTTACACCGGCTGCTCAATGCTGTG 1320
OY      1424  AATCTGTTCCGCTGCTGAGCTGTACAGGACGACGAGAGTACAGAGCGCTCTACAG 1483
Db      1321  AATCTGTTCCGCTGCTGAGCTGTACAGGACGACGAGAGTACAGAGCGCTCTATCAT 1380
OY      1484  AAGTGGCAGATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1543
Db      1381  AAGTGGCAGATGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440

```

```

OY      1544  ATGATTTTCAGAGCTGTGCTACTGAGACGCTGGGCTTACATCTGAGTTGCCGATTT 1603
Db      1441  GTCAATTTTCAGAGCTGTGCTACTGAGACGCTGGGCTTATGATCTGAGTTGCCGATTT 1500
OY      1604  GGATATTTTCTGCTGCTCTTCTTGGCCCCACATTAATTTGATGATTTCTTACTCTGTG 1663
Db      1501  GGATATTTTCTGCTGCTCTTCTTGGCCCCACATTAATTTGATGATTTCTTACTCTGTG 1560
OY      1664  CTACTTGGTATCTGCCAAATTCAAATATGTACAGAGTATGCTGCTGCTGCTCAT 1723
Db      1561  CTGCTTGGTATATCTGCAAAACCTTAATATGTATGATGATGATGATGATGATGATGAT 1620
OY      1724  GCGGGGCTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1783
Db      1621  TCTGGGCTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
OY      1784  AAAATCATCAGTTATTTTACATTCACAAATATTTGAGAGATTTCTTGAATCAATGAG 1843
Db      1681  AAAATCTGGGATTTATTTACATTCACAAATATTTGAGAGATTTCTTGAATCAATGAG 1740
OY      1844  TTTGAGGAGCTGATTTTACATTTGAGAGCTCAAAATGTTTCTGTACAACTAAATGCAATG 1903
Db      1741  TTTTACGGGCTTGAATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800
OY      1904  TGTGCTTCACTGAGGAATTCATTCATGAGAAACCTGCCAGAGTGCATATCTAGA 1963
Db      1801  TGGCCATTCACCCAGAGGGGCTCAGTTATGAGAAACCTGCCAGAGTGCATATCTAGA 1860
OY      1964  TTTCAATGAACTTTGATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2018
Db      1861  TTTACGGCAAACTTCTCTCATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1915

RESULT 10
ABK51687
ID      ABK51687 standard; cdna; 1069 bp.
XX
AC      ABK51687;
XX
DT      30-JUL-2002 (first entry)
XX
DE      cdna encoding hamster ABCG5 protein.
XX
KW      Hamster; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
KW      arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
KW      ss.
XX
OS      Cricetinae sp.
XX
FH      key
FT      CDS
FT      location/Qualifiers
FT      30..1049
FT      /tag= a
FT      /partial
FT      /product= "Hamster ABCG5 protein"
XX      /note= "this sequence lacks both a start and stop codon"
PN      MO200227016-A2.
XX
PD      04-APR-2002.
XX
PF      25-SEP-2001; 2001MO-US29859.
XX
PR      25-SEP-2000; 2000US-235268P.
XX
PA      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA      (PATE/) PATEL S B.
PA      (DEAN/) DEAN M.
XX
PI      Patel SB, Dean M;
XX
DR      WPI; 2002-416483/44.
DR      P-PSDB; AA096987.

```

xx Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
 PT nucleic acid encoding the polypeptide, useful for treating
 PT sitosterolemia, arteriosclerosis and heart diseases
 XX
 XX Example 3; Page 47; 66pp; English.

CC The present invention relates to a new mammalian ATP-binding cassette
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart
 CC disease. The molecules of the invention are also useful for identifying
 CC a compound which alters ABCG5 activity level comprising contacting a cell
 CC culture or mammal which have ABCG5 polypeptide with a compound and
 CC measuring ABCG5 biological activity in the cell culture or in mammal,
 CC where an increase or decrease in ABCG5 biological activity compared to
 CC ABCG5 biological activity in a control cell culture or mammal not
 CC contacted with the compound, identifies a compound that increases or
 CC decreases ABCG5 activity respectively. The cell culture or mammal
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
 CC polypeptide in a cell culture or mammal is also compared with that of a
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
 CC Stimulation of ABCG5 activity is useful for treating or preventing
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
 CC disease. The method of the invention is useful for increasing cholesterol
 CC excretion and/or decreasing cholesterol adsorption. The present nucleic
 CC acid sequence encodes the hamster ABCG5 protein of the invention.
 CC
 XX

Sequence 1069 BP; 266 A; 282 C; 273 G; 248 T; 0 other;

Query Match 32.9%; Score 770.2; DB 24; Length 1069;
 Best Local Similarity 83.7%; Pred. No. 2.2e-203;
 Matches 896; Conservative 0; Mismatches 173; Indels 2; Gaps 2;

QY 368 TCAGGCTCCGGGAAACACGCTGCTGAGCCATGCTGGGAGGCGGGCGCGCGG 427
 DB 1 TCAGGCTCCGGGAAACACGCTGCTGAGCCATGCTGGGAGGCGGGCGCGCGG 59
 QY 428 ACCTTCTGGGGAGGTGTATGTGAACGCGCGCGCGCGCGCGCGCGCGCGCG 487
 DB 60 ACCCTGGAGGGGAGGTGTGTGAACGCGCGCGCGCGCGCGCGCGCGCGCG 119
 QY 488 TGGTTCCTCTACTCTGCGAGAGGACACCTCTGAGACGCTGACCGTCCGAGAGC 547
 DB 120 TGGTTCCTCTACTCTGCGAGAGGACACCTCTGAGACGCTGACCGTCCGAGAGC 179
 QY 548 CTCACATACACCGCGCTGCGACATCCCGCGCGCGCGCGCGCGCGCGCGCG 607
 DB 180 CTCGCTACACCGCGCTGCGACATCCCGCGCGCGCGCGCGCGCGCGCGCG 239
 QY 608 GTGAGAGCCCTCATGCGAGAGCTGAGTGTGAGCCATGTGCGACAGCAGTATGGCAAC 667
 DB 240 GTAGAGGAGCTGTGAAGAGCTAAGTGTGAGCCAGTGTGCGACAGCAGTATGGCAAC 299
 QY 668 TACAGTGGGGGCAATTTTCACGGGTGAGCGCGCGCGCGCGCGCGCGCGCG 727
 DB 300 TATTAATTTGGGGGCAATTTTCACGGGTGAGCGCGCGCGCGCGCGCGCGCG 359
 QY 728 CTCGAGATCTTAAGTGTGAGTGTGAGCCATGTGCGACAGCAGTATGGCAAC 787
 DB 360 ATTACAGAGCCCAAGATCATGATGTGAGCCCAAGCAGTGTGCGACAGCAGTATGGCAAC 419
 QY 788 GCTAATCATATGCTCTCTCTCTGAGTGTGAGCCATGTGCGACAGCAGTATGGCAAC 847
 DB 420 GCAATATCAATGCTCTCTCTCTGAGTGTGAGCCATGTGCGACAGCAGTATGGCAAC 479
 QY 848 ACCTATTCACAGAGCCCGCTCTGAGTGTGAGCCATGTGCGACAGCAGTATGGCAAC 907
 DB 480 ACCATTCACAGAGCCCGCTCTGAGTGTGAGCCATGTGCGACAGCAGTATGGCAAC 539
 QY 908 TTGCGAGAGCTGATTTCTGTGCGACAGCGGGAATCTGATTTCTTCAATGACTCC 967
 DB 540 TACGGAGAGATGTGTTCTGTGCGACAGCGGGAATCTGATTTCTTCAATGACTCT 599

QY 968 GATTACCTCTCTCTGAGTGTGAGCCATGTGCGACAGCAGTATGGCAAC 1027
 DB 600 GATTACCTCTCTCTGAGTGTGAGCCATGTGCGACAGCAGTATGGCAAC 659
 QY 1028 GATACCCCAAGAGGAAATGAGAACTTCAAGAGAGTCCAGATGATGAAATCT 1087
 DB 660 GATACCCCAAGAGGAAATGAGAACTTCAAGAGAGTCCAGATGATGAAATCT 719
 QY 1088 GCCATCAAGAAATGAGCAATTTGATTAATTTGAAATATTGAAATGAAATGAAACAC 1147
 DB 720 GCTTCAGAGAGCTGTGAGTGTGCAAAAATCTCGGAGATATTGAAAGAGCAAAACAC 779
 QY 1148 CTGAAGAGTGAATGATGTTCTTCAAAAACCAAGATCTCTGAGATTTCTCAAA 1207
 DB 780 CTGAAGAGTGAATGATGTTCTTCAAAAACCAAGATCTCTGAGATTTCTCAAA 839
 QY 1208 CTGGGTGTCTCTGAGGAGAGTGAACAAGAACTTGGTGAAGATTAAGCTGAGTAT 1267
 DB 840 CTGGGTGTCTCTGAGGAGAGTGAACAAGAACTTGGTGAAGATTAAGCTGAGTAT 899
 QY 1268 ACGGCTCTCTCTGAGGAGAGTGAACAAGAACTTGGTGAAGATTAAGCTGAGTAT 1327
 DB 900 ATGGCTCTCTCTGAGGAGAGTGAACAAGAACTTGGTGAAGATTAAGCTGAGTAT 959
 QY 1328 CGAAGCAATGCTGAAGAGGCTGATTCAGAGAGCGCTGATGCTTACAGATTTGTC 1387
 DB 960 CAGAGCAATGCTGAAGAGGCTGATTCAGAGAGCGCTGATGCTTACAGATTTGTC 1018
 QY 1388 GGCCTCCACCCCTGATACAGAGCAGTGAAGCTGATGCTTACAGATTTGTC 1438
 DB 1019 GCGGCCACCCCTGATACAGAGCAGTGAAGCTGATGCTTACAGATTTGTC 1069

RESULT 11
 AAD22022
 ID AAD22022 standard; DNA; 472 BP.
 XX
 AC AAD22022;
 DT 12-FEB-2002 (first entry)
 XX
 XX Human sitosterolemia susceptibility gene (SSG) exon 13.
 DE
 XX Human; sitosterolemia susceptibility gene; SSG; arteriosclerosis;
 KW sterol-related disorder; hyperlipidaemia; hypercholesterolemia; therapy;
 KW gall stone; coronary heart disease; cardiovascular disease; arthritis;
 KW xanthoma; haemolytic anaemia; transgenic animal; ds.
 XX
 OS Homo sapiens.
 XX
 PN MO200179272-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 18-APR-2001; 2001MO-US12758.
 XX
 PR 18-APR-2000; 2000US-198465P.
 PR 15-MAY-2000; 2000US-204234P.
 XX
 PA (TULA-) TULARIK INC.
 XX
 PI Tian H, Schultz J, Shan B;
 DR WPI; 2002-017598/02.
 XX
 PT Novel sitosterolemia susceptibility gene polypeptide and
 PT polynucleotide, useful for screening a compound that increases the
 PT level of expression or activity of SSG polypeptide for treating
 PT sterol-related disorder
 PS
 XX Claim 73; Fig 14B; 105pp; English.

CC The invention relates to an isolated sitosterolaemia susceptibility gene
 CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
 CC binding cassette (ABC) family cholesterol transporter. SSG is useful
 CC for identifying a compound useful in the treatment or prevention of a
 CC sterol-related disorder, including sitosterolaemia, hyperlipidaemia,
 CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or
 CC nutritional deficiencies. SSG is also useful for treating cholesterol-
 CC associated diseases or conditions including coronary heart disease and
 CC other cardiovascular diseases, and sitosterolaemia-associated condition
 CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
 CC expression cassette is useful in the production of transgenic non-human
 CC animals. SSG genes and their homologues are useful as tools for a number
 CC of applications including diagnosing sitosterolaemia and other
 CC cardiovascular disorders, for forensics and paternity determinations,
 CC and for treating any of a large number of SSG associated diseases. The
 CC present sequence is an exon of human SSG DNA.

XX Sequence 472 BP: 134 A; 93 C; 100 G; 145 T; 0 other;

Query Match 20.2%; Score 472; DB 24; Length 472;
 Best Local Similarity 100.0%; Pred. No. 1.2e-120;
 Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1869 GCAGCTCAATGTTTGTGACAACTAATCCATGTGCTTCACTCAAGAAATTCAT 1928
 DB 1 GCAGCTCAATGTTTGTGACAACTAATCCATGTGCTTCACTCAAGAAATTCAT 60
 OY 1929 TCATTGAGAAAACCTGCCAGGTGACAACTATGATTCACAAATGAACTTTCTATTG 1988
 DB 61 TCATTGAGAAAACCTGCCAGGTGACAACTATGATTCACAAATGAACTTTCTATTG 120
 OY 1989 ATTGATTTTTCAGCTCTTGTTCATCTTGAAGAAATGTTGTTTCAAAATAGGATC 2048
 DB 121 ATTGATTTTTCAGCTCTTGTTCATCTTGAAGAAATGTTGTTTCAAAATAGGATC 180
 OY 2049 TCATTGACGAGTAGTGAAGACATGGCTGGGAAAAGAAAGCAAGCACTGCCACTGTGCA 2108
 DB 181 TCATTGACGAGTAGTGAAGACATGGCTGGGAAAAGAAAGCAAGCACTGCCACTGTGCA 240
 OY 2109 TCAGCTGCTGGAACGCTGTAAGATGAGAGTGCATGATTTCTTTCTTGACAGACATCTC 2168
 DB 241 TCAGCTGCTGGAACGCTGTAAGATGAGAGTGCATGATTTCTTTCTTGACAGACATCTC 300
 OY 2169 AAGCTTTTAACCATTAACATCCATTTGTGCTTGTGATCCAGACAGCCTTGATGC 2228
 DB 301 AAGCTTTTAACCATTAACATCCATTTGTGCTTGTGATCCAGACAGCCTTGATGC 360
 OY 2229 AATGGAAGTGTATTATAGTCCCTTGCTTACAACTGAGGAGACATGCTATTGCA 2288
 DB 361 AATGGAAGTGTATTATAGTCCCTTGCTTACAACTGAGGAGACATGCTATTGCA 420
 OY 2289 AATTGTGACTGAGCGGACCAAGAAATGTAATTAATTCATTAACCTATGGG 2340
 DB 421 AATTGTGACTGAGCGGACCAAGAAATGTAATTAATTCATTAACCTATGGG 472

RESULT 12
 AAC76065
 ID AAC76065 standard; cDNA: 432 BP.

XX AAC76065;

XX 08-FEB-2001 (first entry)

DE Human ORFX ORF1620 polynucleotide sequence SEQ ID NO:1239.

XX Human: open reading frame; ORFX: detection; cytoskeletal; hepatotropic;
 KW vulnery; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antirheumatic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antipneumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;

KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.

OS Homo sapiens.

PN W0200058473-A2.

PD 05-OCT-2000.

PF 31-MAR-2000; 2000MO-US08621.

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

PA (CURA-) CURAGEN CORP.

PI Shinkets RA, Leach M;

DR WPI; 2000-602362/57.

XX P-PDB; AAB41856.

PS Novel nucleic acids and peptides derived from open reading frame X,
 useful for treating e.g. cancers, proliferative disorders,
 neurodegenerative disorders and cardiovascular disease -
 claim 5; Page 2444; 5507pp; English.

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytoskeletal; hepatotropic; vulnery;
 CC antiparkinsonian; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antirheumatic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antinflammatory; antibacterial; antiviral; antifungal; antipneumatic;
 CC antithyroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 432 BP: 87 A; 110 C; 118 G; 117 T; 0 other;

Query Match 18.4%; Score 429.4; DB 21; Length 432;
 Best Local Similarity 99.8%; Pred. No. 7.9e-109;
 Matches 430; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1175 AAACCAAGATTCCTCGAGAGTTTCTTAACCTGAGGTGTTCTCTGAGAGAGTGACA 1234
 DB 1 AAACCAAGATTCCTCGAGAGTTTCTTAACCTGAGGTGTTCTCTGAGAGAGTGACA 60
 OY 1235 AGAACTTGAGAAATAGCTGCACTGATAGCGCTCCCTTCAAGATCTGATG 1294
 DB 61 AGAACTTGAGAAATAGCTGCACTGATAGCGCTCCCTTCAAGATCTGATG 120
 OY 1295 GGTGTTGCTCTCTTTCTTCTGTTCTGCGGCTCGAAGCAATGCTTAAGGCTGCTATC 1354
 DB 121 GGTGTTGCTCTCTTTCTTCTGTTCTGCGGCTCGAAGCAATGCTTAAGGCTGCTATC 180

QY 1355 CAGGACCGCGTAGGTCCTTTACGAGTTGTGGGCGCACCCCGTACACAGGCATGCTG 1414
DB 181 CAGGACCGCGTAGGTCCTTTACGAGTTGTGGGCGCACCCCGTACACAGGCATGCTG 240
QY 1415 AACGCTGTGAATGTGTTCCCGTGCAGAGCTGTGTAGCAGCAGAGAGTCAGAGCGC 1474
DB 241 AACGCTGTGAATGTGTTCCCGTGCAGAGCTGTGTAGCAGCAGAGAGTCAGAGCGC 300
QY 1475 CTCCTACAGAGTAGGAGATGATGCTGCTATGACAGTCAGCTCTCCCTTACAGCCTT 1534
DB 301 CTCCTACAGAGTAGGAGATGATGCTGCTATGACAGTCAGCTCTCCCTTACAGCCTT 360
QY 1535 GTTGGCAGCATGATTTTTCAGCAGTGTGTCTACTAGCAGCCTGGGCTTACATCCTGAGTT 1594
DB 361 GTTGGCAGCATGATTTTTCAGCAGTGTGTCTACTAGCAGCCTGGGCTTACATCCTGAGTT 420
QY 1595 GCCCGATTGGG 1605
DB 421 GCCCGATTGGG 431

RESULT 13

AA294755
ID AA294755 standard; cDNA; 281 BP.

XX AA294755;

DT 01-AUG-2000 (first entry)

DE Human ATP binding cassette cDNA fragment 168043.

XX ATP binding cassette; human; cholesterol; lipid disorder;
KW atherosclerosis; lipid disorder; dyslipidemia; psoriasis;
KM lupus erythematosus; diagnosis; gene therapy; ss.
XX

OS Homo sapiens.

XX WO200018912-A2.

PD 06-APR-2000.

PF 21-SEP-1999; 99WO-EP06991.

PR 25-SEP-1998; 98US-0101706.

PA (FARB) BAYER AG.

PI Schmitz G, Klucken J;

DR WPI; 2000-293151/25.

PT Adenosine triphosphate binding proteins useful for identifying agents
for treating atherosclerosis and other inflammatory disorders -
XX Claim 9; Page 135; 154pp; English.CC The present sequence is that of human ATP binding cassette
CC (ABC) cDNA fragment 168043, identified as a cholesterol-sensitive
CC gene fragment. The invention provides cholesterol-sensitive ABC
CC genes (see AA294734-63). These genes, and polypeptides encoded by
CC them, can be used for diagnostic and therapeutic applications, and
CC for biochemical or cell-based assays to screen for pharmacologically
CC active modulator compounds useful for the treatment of lipid
CC disorders, atherosclerosis or other inflammatory diseases such as
CC psoriasis and lupus erythematosus.
XX

SQ Sequence 281 BP; 60 A; 68 C; 73 G; 80 T; 0 other;

Query Match 11.5%; Score 268; DB 21; Length 281;
Best Local Similarity 99.6%; Pred. No. 4.8e-64;
Matches 279; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1175 AAACCAAGATTTCTCCGTGAGTTTCTCTAAACGTGGTGTCTCTGAGAGAG--TGAC 1233
DB 1 AAACCAAGATTTCTCCGTGAGTTTCTCTAAACGTGGTGTCTCTGAGAGAGTTGAC 60
QY 1234 AAGAACTTGATGAGAAATTAAGCTGACAGTGTATACGCTCTCAGAACTGATCAT 1293
DB 61 AAGAACTTGATGAGAAATTAAGCTGACAGTGTATACGCTCTCAGAACTGATCAT 120
QY 1294 GGGTTGTTCTCTCTTTTCTTTGTTGCTGCGGGTCCGAACAATGCTAAAGGTGCTAT 1353
DB 121 GGGTTGTTCTCTCTTTTCTTTGTTGCTGCGGGTCCGAACAATGCTAAAGGTGCTAT 180
QY 1354 CCAGACCGCGTAGGTCCTCTTACAGATTGTGGGCGCGCACCCCGTACAGAGCATGCT 1413
DB 181 CCAGACCGCGTAGGTCCTCTTACAGATTGTGGGCGCGCACCCCGTACAGAGCATGCT 240
QY 1414 GAACGCTGTGAATCTGTTTCCCGTGTGCGAGCCTGTACAC 1453
DB 241 GAACGCTGTGAATCTGTTTCCCGTGTGCGAGCTGTACAC 280

RESULT 14

ABK51683
ID ABK51683 standard; DNA; 5460 BP.

XX ABK51683;

DT 30-JUL-2002 (first entry)

DE Human ABCG5 upstream genomic sequence, exon 1, intron 1 and exon 2.

XX Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;
KW atherosclerosis; heart disease; hypersterolemia; Alzheimer's disease;
KM chromosome 2p21; ds.
XX

OS Homo sapiens.

XX WO200227016-A2.

PD 04-APR-2002.

PF 25-SEP-2001; 2001WO-US29859.

PR 25-SEP-2000; 2000US-235268P.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

PI Patel SB, Dean M;

DR WPI; 2002-416483/44.

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the
nucleic acid encoding the polypeptide, useful for treating
XX sitosterolemia, atherosclerosis and heart diseases -
XX Example 3; Page 38-41; 66pp; English.CC The present invention relates to a new mammalian ATP-binding cassette
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a
CC predisposition for developing sitosterolemia, atherosclerosis or heart
CC disease. The molecules of the invention are also useful for identifying
CC a compound which alters ABCG5 activity level comprising contacting a cell
CC culture or mammal which have ABCG5 polypeptide with a compound and
CC measuring ABCG5 biological activity in the cell culture or in mammal,
CC where an increase or decrease in ABCG5 biological activity compared to
CC ABCG5 biological activity in a control cell culture or mammal not
CC contacted with the compound, identifies a compound that increases or
CC decreases ABCG5 activity respectively. The cell culture or mammal
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the
CC polypeptide in a cell culture or mammal is also compared with that of a

CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.
CC Stimulation of ABCG5 activity is useful for treating or preventing
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's
CC disease. The method of the invention is useful for increasing cholesterol
CC excretion and/or decreasing cholesterol adsorption. The present nucleic
CC acid sequence represents the upstream genomic sequence, exon 1, intron 1
CC and exon 2 of the human ABCG5 gene located on chromosome 2p21.
XX

Sequence 5460 BP; 1351 A; 1350 C; 1508 G; 1243 T; 8 other;

Query Match 10.7%; Score 249.6; DB 24; Length 5460;
Best Local Similarity 98.4%; Pred. No. 3.8e-58;
Matches 252; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 GTCAGGTGGAGCAGGAGGAGTCTGCTCCAGGGGCTCCCACTGAAAGCCACTCTGGGGA 60
DB 4504 GTCAGGTGGAGCAGGAGGAGTCTGCTCCAGGGGCTCCCACTGAAAGCCACTCTGGGGA 4563
OY 61 GGGTCGGGCGCCACAGAAATTTGGCCACTTGGCTGTGGCCATGGGTGACCTCTC 120
DB 4564 GGGTCGGGCGCCACAGAAATTTGGCCACTTGGCTGTGGCCATGGGTGACCTCTC 4623
OY 121 ATCTTTGACCCCGGAGGCTCATGGGTCTCCAAAGTAAACAGAGGCTCCAGAGCTCCCT 180
DB 4624 ATCTTTGACCCCGGAGGCTCATGGGTCTCCAAAGTAAACAGAGGCTCCAGAGCTCCCT 4683
OY 181 GGAGGGGCTCTCTGCTCCAGCCCGCGAGCTCTCAAGCTTGAGCACTCTCCCTCTCTA 240
DB 4684 GGAGGGGCTCTCTGCTCCAGCCCGCGAGCTCTCAAGCTTGAGCACTCTCCCTCTCTA 4743
OY 241 CAGCGTCAGCCAGCC 256
DB 4744 CAGCGTCAGTAAAGCC 4759

RESULT 15

AAD22010
ID AAD22010 standard; DNA; 249 BP.

XX AAD22010;

XX 12-FEB-2002 (first entry)

XX Human sitosterolemia susceptibility gene (SSG) exon 1.

KW Human; sitosterolemia susceptibility gene; SSG; arteriosclerosis;
KW sterol-related disorder; hyperlipidaemia; hypercholesterolemia; therapy;
KW gall stone; coronary heart disease; cardiovascular disease; arthritis;
KW xanthoma; haemolytic anaemia; transgenic animal; ds.

XX Homo sapiens.

XX WO200179272-A2.

XX 25-OCT-2001.

XX 18-APR-2001; 2001WO-US12758.

XX 18-APR-2000; 2000US-198465P.

XX 15-MAY-2000; 2000US-204234P.

XX (TULA-) TULARIK INC.

XX Tian H, Schultz J, Shan B;

XX WPI; 2002-017598/02.

XX Novel sitosterolemia susceptibility gene polypeptide and
XX polynucleotide, useful for screening a compound that increases the
XX level of expression or activity of SSG polypeptide for treating
XX sterol-related disorder

XX Claim 73; Flg 14B; 105pp; English.

XX The invention relates to an isolated sitosterolemia susceptibility gene
CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)
CC binding cassette (ABC) family cholesterol transporter. SSG is useful
CC for identifying a compound useful in the treatment or prevention of a
CC sterol-related disorder, including sitosterolemia, hyperlipidaemia,
CC hypercholesterolemia, gall stones, HDL deficiency, arteriosclerosis or
CC nutritional deficiencies. SSG is also useful for treating cholesterol-
CC associated diseases or conditions including coronary heart disease and
CC other cardiovascular diseases, and sitosterolemia-associated condition
CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG
CC expression cassette is useful in the production of transgenic non-human
CC animals. SSG genes and their homologues are useful as tools for a number
CC of applications including diagnosing sitosterolemia and other
CC cardiovascular disorders, for forensics and paternity determinations,
CC and for treating any of a large number of SSG associated diseases. The
CC present sequence is an exon of human SSG DNA.

Sequence 249 BP; 44 A; 86 C; 74 G; 45 T; 0 other;

Query Match 10.6%; Score 249; DB 24; Length 249;
Best Local Similarity 100.0%; Pred. No. 8.6e-59;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCAGGTGGAGCAGGAGGAGTCTGCTCCAGGGGCTCCCACTGAAAGCCACTCTGGGGA 60
DB 1 GTCAGGTGGAGCAGGAGGAGTCTGCTCCAGGGGCTCCCACTGAAAGCCACTCTGGGGA 60
OY 61 GGGTCGGGCGCCACAGAAATTTGGCCACTTGGCTGTGGCCATGGGTGACCTCTC 120
DB 61 GGGTCGGGCGCCACAGAAATTTGGCCACTTGGCTGTGGCCATGGGTGACCTCTC 120
OY 121 ATCTTTGACCCCGGAGGCTCATGGGTCTCCAAAGTAAACAGAGGCTCCAGAGCTCCCT 180
DB 121 ATCTTTGACCCCGGAGGCTCATGGGTCTCCAAAGTAAACAGAGGCTCCAGAGCTCCCT 180
OY 181 GGAGGGGCTCTCTGCTCCAGCCCGCGAGCTCTCAAGCTTGAGCACTCTCCCTCTCTA 240
DB 181 GGAGGGGCTCTCTGCTCCAGCCCGCGAGCTCTCAAGCTTGAGCACTCTCCCTCTCTA 240
OY 241 CAGCGTCAG 249
DB 241 CAGCGTCAG 249

Search completed: July 27, 2003, 00:55:32
Job time : 570.197 secs

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

(without alignments)

11512.153 Million cell updates/sec

US-09-989-981A-5

2340

“

IDENTITY_NUC

4
1
.
P
:
:
:

length: 0

Minimum Match 0%

Maximum Match 100%

6444

```

1:  gb.ba:*
2:  gb.htg:*
3:  gb.in:*
4:  gb.om:*
5:  gb.ov:*
6:  gb.pat:*
7:  gb.ph:*
8:  gb.pl:*
9:  gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.ov:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*
29: em.vi:*
30: em.hg.hum:*
31: em.hg.in:*
32: em.hg.other:*
33: em.hg.mus:*
34: em.hg.pln:*
35: em.hg.prod:*
36: em.hg.nam:*
37: em.hg.vrt:*
38: em.sv:*
39: em.hgo.hum:*
40: em.hgo.mus:*
41: em.hgo.other:*

```

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB	ID	Description
1	2340	100.0	2340	6	AX320883	AX320883 Sequence
2	2340	100.0	2340	6	AX685733	AX685733 Sequence
3	2340	100.0	2340	9	AF320293	AF320293 Homo sapl
4	2340	100.0	2740	9	AF312715	AF312715 Homo sapl
5	2338.4	99.9	2510	6	AX455520	AX455520 Sequence
6	1920	82.1	1920	6	AX456519	AX456519 Sequence
7	1410.8	60.3	2251	10	AY195873	AY195873 Mus muscu
8	1409.2	60.2	2354	6	AX456524	AX456524 Sequence
9	1409.2	60.2	2354	10	AF312713	AF312713 Mus muscu
10	1406	60.1	2351	10	AY195872	AY195872 Mus muscu
11	1395.6	59.6	2258	6	AX320881	AX320881 Sequence
12	1383.8	59.1	2470	10	AF312714	AF312714 Rattus no
13	1365.4	58.4	1959	6	AX685729	AX685729 Sequence
14	1363	58.2	2035	6	AX456526	AX456526 Sequence
15	1335.8	57.1	1915	6	AX456523	AX456523 Sequence
16	1174.2	50.2	2512	9	AK091997	AK091997 Homo sapl
17	770.2	32.9	1069	6	AX011528	AX455528 Sequence
18	473	20.2	186091	9	AC011242	AC011242 Homo sapl
19	472	20.3	472	6	AX320898	AX320898 Sequence
20	312	13.3	127066	9	AC084265	AC084265 Homo sapl
21	312	13.3	139342	9	AC108476	AC108476 Homo sapl
22	249.6	10.7	2809	9	F351812501	F351812501 Homo sapl
23	249.6	10.7	4899	9	AF404106	AF404106 Homo sapl
24	249.6	10.7	5459	6	AX455521	AX455521 Sequence
25	249	10.6	249	6	AX320885	AX320885 Sequence
26	249	10.6	581	9	AF404107	AF404107 Homo sapl
27	230.2	9.8	66166	2	AC084712	AC084712 Homo sapl
28	214	9.1	214	6	AX320893	AX320893 Sequence
29	206	8.8	206	6	AX320894	AX320894 Sequence
30	203.6	8.7	2022	9	AF320292	AF320292 Homo sapl
31	203.6	8.7	2669	6	AX685735	AX685735 Sequence
32	203.6	8.7	2679	9	AF324494	AF324494 Homo sapl
33	200.8	8.6	2285	10	AY196215	AY196215 Mus muscu
34	199.2	8.5	2019	6	AX685731	AX685731 Sequence
35	199.2	8.5	2284	10	AY196216	AY196216 Mus muscu
36	189.2	8.5	3674	10	AF324495	AF324495 Mus muscu
37	189.2	8.1	4829	10	AF311785	AF311785 Rattus no
38	186	7.9	186	6	AX320896	AX320896 Sequence
39	176.2	7.5	40929	10	AY185899	AY185899 Rattus no
40	174.6	7.5	237445	2	AC120701	AC120701 Rattus no
41	174.6	7.5	312858	2	AC112747	AC112747 Rattus no
42	171.2	7.3	92535	2	AC093340	AC093340 Rattus no
43	166	7.1	175439	8	AP001111	AP001111 Oryzola
44	157.6	6.7	165113	8	AC084405	AC084405 Oryza sat
45	153.6	6.6	175439	8	AP001111	AP001111 Oryza sat

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
AX320883	AX320883	Sequence 4 from Patent WO0179272.	AX320883	AX320883.1	GI:17902433	Homo sapiens (human)	Homo sapiens	1	Tian, H., Schultz, J. and Shan, B.	Sitosterolemia susceptibility gene (ssg): compositions and methods of use

JOURNAL Patent: WO 0179272-A 4 25-OCT-2001;
Tularik Inc. (US)
FEATURES Location/Qualifiers
Source 1. 2340

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="human stitosterolemia gene (SSG)"
107..2062
CDS
/note="human stitosterolemia susceptibility gene (SSG)
protein"
/protein_id="CAD19409.1"
/db_xref="GI:1792434"

/translation="MGDLSLTPGSGMGLQVNGSOSLSLEGAPATAPEPHSLGILHAS
YSVSHVRPMMDITSCROOITROILKDVSLYVESGOIMCILGSSGGKTLIDAMSGR
LGRAGTFLFEVYNGRRLRREPODFEVLSDTLSSITVEETHYALALIRGN
PGSFORKVEVNAVELSLSHVADRLIGNVSLGGISTGERRVSIAALIDDPKMLEDE
PAMLDPEKDCGTPCEHSPNPEFMDLTSVDPQSKEREIETSKRVOMTESAVKRSKI
CHRTLNIEBMRHKLTPMPVPEKTPSPVSKLGLRLRRVRLNRLNLAIVTLILLO
NLIMGLFEFLYLRVSNVLKGIODRGLIYQVATPYTGMLNVLNLEPVLRAVSD
OESODGLYOKROMMLAYALHVLPEFVYATMIFSSVCYMTGLHPEVARFGYSALILA
PHLIGELTLVLIGIVONPNVNSVALLSINGVILVSGFPIRIQEMPIPRKISITFT
FOKYSSEILVNVNFEVGLNFTCCSSNVSVTTPNMCATGCIOTIEKTCPEATSRFTNMF
LIIKSTPIALVILIEIVETKIRDLISR"

BASE COUNT 541 a 601 c 598 g 600 t
ORIGIN

Query Match 100.0% Score 2340; DB 6; Length 2340;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCAGGTGAGCAGCAGGAGCTGCGACAGGGCTCCCAACTGAAGCACTCTGGGG 60
DB 1 GTCAGGTGAGCAGCAGGAGCTGCGACAGGGCTCCCAACTGAAGCACTCTGGGG 60
OY 61 GGGTCCGGCCACAGAAAATTTGCCAGCTTTCCTGCTGGCCATGGGTGACCTCTC 120
DB 61 GGGTCCGGCCACAGAAAATTTGCCAGCTTTCCTGCTGGCCATGGGTGACCTCTC 120
OY 121 ATCTTGACCCCGGAGGGGTCATGGTCTCCAGTAAGAACAGAGCTCCCAAGCTCCCT 180
DB 121 ATCTTGACCCCGGAGGGGTCATGGTCTCCAGTAAGAACAGAGCTCCCAAGCTCCCT 180
OY 122 ATCTTGACCCCGGAGGGGTCATGGTCTCCAGTAAGAACAGAGCTCCCAAGCTCCCT 180
DB 122 ATCTTGACCCCGGAGGGGTCATGGTCTCCAGTAAGAACAGAGCTCCCAAGCTCCCT 180
OY 181 GAGAGGGGCTCTGCGACCGCCCGGAGGCTCACAAGCTGAGGCTCATCTCATGCTCTA 240
DB 181 GAGAGGGGCTCTGCGACCGCCCGGAGGCTCACAAGCTGAGGCTCATCTCATGCTCTA 240
OY 241 CAGCTCAGCAGCAGCAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 300
DB 241 CAGCTCAGCAGCAGCAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 300
OY 301 CAGCAGATCTCAAGATGCTCTGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
DB 301 CAGCAGATCTCAAGATGCTCTGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
OY 361 AGGAACTCAGGCTCCGGGAAAACACAGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 420
DB 361 AGGAACTCAGGCTCCGGGAAAACACAGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 420
OY 421 CGGGGAGACCTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480
DB 421 CGGGGAGACCTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480
OY 481 CCAGAGCTCTTCTCTACGCTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
DB 481 CCAGAGCTCTTCTCTACGCTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
OY 541 CGAGAGCTGCACTACACCGGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600
DB 541 CGAGAGCTGCACTACACCGGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600

OY 601 GAAGAGGTGAGAGCCGCTCATGAGAGGCTGAGTGCATGTGCGACAGCCAGCTGAT 660
DB 601 GAAGAGGTGAGAGCCGCTCATGAGAGGCTGAGTGCATGTGCGACAGCCAGCTGAT 660
OY 661 TGGCAACTACAGCTTGGGGGGCATTTTCACAGGGTGAGCGCGCGGCTCCATCCGAC 720
DB 661 TGGCAACTACAGCTTGGGGGGCATTTTCACAGGGTGAGCGCGCGGCTCCATCCGAC 720
OY 721 CCAGCTGCTCCAGAGCTCAAGTCAATGCTGTTGATGAGCCACACAGGAGGCTGACAG 780
DB 721 CCAGCTGCTCCAGAGCTCAAGTCAATGCTGTTGATGAGCCACACAGGAGGCTGACAG 780
OY 781 CATGAGCTGTAATCAGATTTGCTGCTCCCTGAGTGAACAGGCTGCGAGAACGAAATTTG 840
DB 781 CATGAGCTGTAATCAGATTTGCTGCTCCCTGAGTGAACAGGCTGCGAGAACGAAATTTG 840
OY 841 GGTTCACACATTCACACAGCCCGGCTGAGCTTTTTCAGCTTTTTCAGAAAATTCAT 900
DB 841 GGTTCACACATTCACACAGCCCGGCTGAGCTTTTTCAGCTTTTTCAGAAAATTCAT 900
OY 901 CCTGAGCTTCGAGAGCTGATTTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 960
DB 901 CCTGAGCTTCGAGAGCTGATTTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 960
OY 961 TGACCTGCGGTACCCCTTGTCTGAGACATTCAAACCCCTTTCATATGAGACCTGAC 1020
DB 961 TGACCTGCGGTACCCCTTGTCTGAGACATTCAAACCCCTTTCATATGAGACCTGAC 1020
OY 1021 GTCAGTGTATACCAAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
DB 1021 GTCAGTGTATACCAAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
OY 1081 AGAATCTGCTCAAGAAATTCAGCAATTTGCTATAAACTTTGAAGAAATTTGAAGAAAT 1140
DB 1081 AGAATCTGCTCAAGAAATTCAGCAATTTGCTATAAACTTTGAAGAAATTTGAAGAAAT 1140
OY 1141 GAAACACCTGAAAGCTTACCAATGCTTCCCTTCAAAACCAAGATTCCTGAGATTT 1200
DB 1141 GAAACACCTGAAAGCTTACCAATGCTTCCCTTCAAAACCAAGATTCCTGAGATTT 1200
OY 1201 CTTCTAATCTGGGCTTCTCTGAGGAGGAGTGAACAGAAACTTGTGAGAAATTAAGCTG 1260
DB 1201 CTTCTAATCTGGGCTTCTCTGAGGAGGAGTGAACAGAAACTTGTGAGAAATTAAGCTG 1260
OY 1261 AGTGAATACGCTTCAAGATCTGATGATGATGATGATGATGATGATGATGATGATGATG 1320
DB 1261 AGTGAATACGCTTCAAGATCTGATGATGATGATGATGATGATGATGATGATGATGATG 1320
OY 1321 GCGGGTCCGAAACAAATGCTAAAGGAGTCTATCCAGAGCAGGAGTCTCTTACCA 1380
DB 1321 GCGGGTCCGAAACAAATGCTAAAGGAGTCTATCCAGAGCAGGAGTCTCTTACCA 1380
OY 1381 GTTGTGGGCGCCACCCGCTACACAGGAGTGAACGCTGTGAATCTGTTCCCGTCT 1440
DB 1381 GTTGTGGGCGCCACCCGCTACACAGGAGTGAACGCTGTGAATCTGTTCCCGTCT 1440
OY 1441 GCGAGCTGACAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
DB 1441 GCGAGCTGACAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
OY 1501 GGCCTATGACAGTCACTCTCCCTTACAGCTTGTTCACACATGATTTTCAGAGAGT 1560
DB 1501 GGCCTATGACAGTCACTCTCCCTTACAGCTTGTTCACACATGATTTTCAGAGAGT 1560
OY 1561 GTGCTACTGAGAGGCTGAGGCTTACATCTGAGAGTCCGAGATTTGATATTTTTCGTC 1620
DB 1561 GTGCTACTGAGAGGCTGAGGCTTACATCTGAGAGTCCGAGATTTGATATTTTTCGTC 1620
OY 1621 TCTTTGGCCCCCACTTAATTTGAGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
DB 1621 TCTTTGGCCCCCACTTAATTTGAGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
OY 1681 AAATCAAAATATGACAAAGTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740

Db 1681 AATTCGAAATATATGCTAAGAGTGTAGGCTCTGCTCTCCATTCGCGGGGCTGCTTGG 1740

Qy 1741 ATCTGGATTCCTCGAAGAACATACAGAAATGCCCATCTTTTAAATCAATCACTTATTT 1800

Db 1741 ATCTGGATTCCTCGAAGAACATACAGAAATGCCCATCTTTTAAATCAATCACTTATTT 1800

Qy 1801 TACATTCGCAAAATATATGAGTGTAGTCTTGTAGTCAATGAGTCTACGAGCTGAATTT 1860

Db 1801 TACATTCGCAAAATATATGAGTGTAGTCTTGTAGTCAATGAGTCTACGAGCTGAATTT 1860

Qy 1861 CACTGTGCGAGCTCAAAATGTTTCTGTGACAACTAAATCCATGTGCTCTTCACTCAAG 1920

Db 1861 CACTGTGCGAGCTCAAAATGTTTCTGTGACAACTAAATCCATGTGCTCTTCACTCAAG 1920

Qy 1921 AATTCATTCATTCAGAAAGCTGCGCGAGTGCAGAACTAGATTCATGAACTTCT 1980

Db 1921 AATTCATTCATTCAGAAAGCTGCGCGAGTGCAGAACTAGATTCATGAACTTCT 1980

Qy 1981 GATTTTGTATTCATTTATTCACAGCTCTTGTATCTATCTAGAAATGTTTCTTCAAAATAG 2040

Db 1981 GATTTTGTATTCATTTATTCACAGCTCTTGTATCTATCTAGAAATGTTTCTTCAAAATAG 2040

Qy 2041 GATTCATTCATTCAGAGTGTAGTGAAGCCATGCTGCGAAATGGAAGTGAAGCTCCG 2100

Db 2041 GATTCATTCATTCAGAGTGTAGTGAAGCCATGCTGCGAAATGGAAGTGAAGCTCCG 2100

Qy 2101 ACATGTGATGCTGCTGTGAAGCTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2160

Db 2101 ACATGTGATGCTGCTGTGAAGCTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2160

Qy 2161 GACATCTCAAGTCTTTTAAACATTAAAGTCAATTTGTGCTCTTGTGATTCAGAGAGCC 2220

Db 2161 GACATCTCAAGTCTTTTAAACATTAAAGTCAATTTGTGCTCTTGTGATTCAGAGAGCC 2220

Qy 2221 TTTAAATGCAATGCAAGTGTATTTAGTCCCTGCTCTTACACTGCGAGGAGATGTGT 2280

Db 2221 TTTAAATGCAATGCAAGTGTATTTAGTCCCTGCTCTTACACTGCGAGGAGATGTGT 2280

Qy 2281 TATTTGGAATTTGATGAGTGAAGGAGCCCAAGATGTAATTAATTCATTAACCTATGG 2340

Db 2281 TATTTGGAATTTGATGAGTGAAGGAGCCCAAGATGTAATTAATTCATTAACCTATGG 2340

RESULT 2
AX685733 2340 bp DNA linear PAT 29-MAR-2003
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
CDS
FEATURES
source
1. 2340
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
107..2062
/note="unnamed protein product; human ABCG5 (hABCG5)"
/protein_id="CAD86572.1"
/db_xref="GI:29371743"
/translation="MGDLSTLPKGSMSGLQVNRGSSSLGAPATAPBPHSLGILHAS
YSVSHRVPMDITSCROOWTRQILKIDVSLVYESQIMCITLSSSGSKTTLDAASGR

BASE COUNT 541 a 601 c 598 g 600 t
ORIGIN
LGRAGTFLGEVYNGRALREOFDPCSVYIADSLTSLTVEETLHTALALRGN
PGSPORKVEAVMAELSLSHADRLNIGYSLAGSISTGERRRVSIAQIOLDPKWLEDE
PTTGIDCMRANQIVLVELLARRNRIVLTIHQPRELEFLFKIALISGELIFGCT
PAEMLEFNDGYPCEPHSNRPFDEYNDLTSVDPQSEKRELETESRVMESAYKSAI
CHRTKNIEMRMLKTLPMVPFKTDSGPVFSKGLVLRVTNRNLVNRLEAVTRILLO
NLJNGJLLEFVLRVRSNVLKGAIOBPVGLLOFVGATPYTGMLNVNLPVLRAYS
OESODGLYOKOMALAYALHVLPEPSVATWLESVCYRTGLPEVAREYFENALIA
PHLIGETFLVILGIONPNIVNSVALLSIAVLVSGSLRNIOBMPIFPKLITFT
FOYCSSEIIVNERTGIAFTGSSNSVSTTNPMCAFTGIGIOTLEKCPGATSRFTNMF
LIIYSFIPALVILGIYVFKIRDLHSR"

Query Match 100.0%; Score 2340; DB 6; Length 2340;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCAGGTGAGAGGAGGAGGAGCTGTCACAGGAGCTCCCACTGAAAGCCACTGTGGGA 60

Db 1 GTCAGGTGAGAGGAGGAGGAGCTGTCACAGGAGCTCCCACTGAAAGCCACTGTGGGA 60

Qy 61 GGGTCCGGCCACAGAAATTTGCCCAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120

Db 61 GGGTCCGGCCACAGAAATTTGCCCAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120

Qy 121 ATCTTTGACCCCGGAGAGGCTGATGCTGCTCAAGTAAACAGAGCTCCAGAGCTCCT 180

Db 121 ATCTTTGACCCCGGAGAGGCTGATGCTGCTCAAGTAAACAGAGCTCCAGAGCTCCT 180

Qy 181 GAGAGGGGCTCTGTCACAGCCCGGAGCTGACAGCTGGGATGCTGCTGCTGCTGCTGCT 240

Db 181 GAGAGGGGCTCTGTCACAGCCCGGAGCTGACAGCTGGGATGCTGCTGCTGCTGCTGCT 240

Qy 241 CAGCGTCAAGCAGCCGCTGAGGCTGCTGAGGATGATCATCTTGGCGGAGCAGTGTGAC 300

Db 241 CAGCGTCAAGCAGCCGCTGAGGCTGCTGAGGATGATCATCTTGGCGGAGCAGTGTGAC 300

Qy 301 CAGGAGATGCTCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360

Db 301 CAGGAGATGCTCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360

Qy 361 AGGAGACTGAGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420

Db 361 AGGAGACTGAGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420

Qy 421 CGCGGGAGACCTTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 480

Db 421 CGCGGGAGACCTTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 480

Qy 481 CCAGGAGTGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540

Db 481 CCAGGAGTGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540

Qy 541 CGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600

Db 541 CGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600

Qy 601 GAGAGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660

Db 601 GAGAGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660

Qy 661 TGGCAACTACAGCTTGGGGGAGATTTCCAGGGGTGAGCGGGGCGGGTCTCATCCAGC 720

Db 661 TGGCAACTACAGCTTGGGGGAGATTTCCAGGGGTGAGCGGGGCGGGTCTCATCCAGC 720

Qy 721 CCAGCTGCTCCAGGAGCTAAGTCTAGTCTGTTTATGAGCCACAGGCTGAGCTG 780

Db 721 CCAGCTGCTCCAGGAGCTAAGTCTAGTCTGTTTATGAGCCACAGGCTGAGCTG 780

Qy 781 CATGACTGCTAATCAGATTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840

Db 781 CATGACTGCTAATCAGATTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840

OY 841 GGTTCACCAATTCACACGCCGCTTCTGAGCTTTTTCAGCTCTTGAGCAAAATTCGCAT 900
 DB 841 GGTTCACCAATTCACACGCCGCTTCTGAGCTTTTTCAGCTCTTGAGCAAAATTCGCAT 900
 OY 901 CCGAGCTTGGAGAGCTGATTTTCTGTGGCAGCCGAGGGAATGCTGATTTCTTCA 960
 DB 901 CCGAGCTTGGAGAGCTGATTTTCTGTGGCAGCCGAGGGAATGCTGATTTCTTCA 960
 OY 961 TGACCTGGTTACCCCTTGTGCTGCAATTCACACCTTTTGTGCTTATATGAGCTGAC 1020
 DB 961 TGACCTGGTTACCCCTTGTGCTGCAATTCACACCTTTTGTGCTTATATGAGCTGAC 1020
 OY 1021 GTGAGTGTATACCAAGCAAGGAGGGAATAGAAACCTCCAAAGAGTCCAGATGAT 1080
 DB 1021 GTGAGTGTATACCAAGCAAGGAGGGAATAGAAACCTCCAAAGAGTCCAGATGAT 1080
 OY 1081 AGAATCTGCTACAGAAATTCAGCAATTTGTCTAATAAATTGAAATATGAAAGAT 1140
 DB 1081 AGAATCTGCTACAGAAATTCAGCAATTTGTCTAATAAATTGAAATATGAAAGAT 1140
 OY 1141 GAAACACCTGAAAGAGCTTACCAATGCTTCTTCAAAACCAAGATTTCTGAGCTTT 1200
 DB 1141 GAAACACCTGAAAGAGCTTACCAATGCTTCTTCAAAACCAAGATTTCTGAGCTTT 1200
 OY 1201 CTCTAAACTGGGCTTCTCTGAGAGAGTGAACAATAAATTGCTGAGATGCTGCT 1260
 DB 1201 CTCTAAACTGGGCTTCTCTGAGAGAGTGAACAATAAATTGCTGAGATGCTGCT 1260
 OY 1261 AGTGAATACGCTCTCTCTGAGAGAGTGAACAATAAATTGCTGAGATGCTGCT 1320
 DB 1261 AGTGAATACGCTCTCTCTGAGAGAGTGAACAATAAATTGCTGAGATGCTGCT 1320
 OY 1321 GCGGCTCGGAGCAATGCTTAAAGGCTCTATCCAGAGCCGCTGAGTCTCTTACCA 1380
 DB 1321 GCGGCTCGGAGCAATGCTTAAAGGCTCTATCCAGAGCCGCTGAGTCTCTTACCA 1380
 OY 1381 GTTGTGGGCGCCACCCGCTGAGAGAGTGAACAATAAATTGCTGAGATGCTGCT 1440
 DB 1381 GTTGTGGGCGCCACCCGCTGAGAGAGTGAACAATAAATTGCTGAGATGCTGCT 1440
 OY 1441 GCGAGCTGTACAGCAAGAGAGTGAACAATAAATTGCTGAGATGCTGCT 1500
 DB 1441 GCGAGCTGTACAGCAAGAGAGTGAACAATAAATTGCTGAGATGCTGCT 1500
 OY 1501 GCGCTATGACGTGACAGTCTCTCTGAGAGAGTGAACAATAAATTGCTGAGATGCT 1560
 DB 1501 GCGCTATGACGTGACAGTCTCTCTGAGAGAGTGAACAATAAATTGCTGAGATGCT 1560
 OY 1561 GTGCTAGTGAAGAGTGAACAATAAATTGCTGAGAGAGTGAACAATAAATTGCTGAG 1620
 DB 1561 GTGCTAGTGAAGAGTGAACAATAAATTGCTGAGAGAGTGAACAATAAATTGCTGAG 1620
 OY 1621 TCTCTTGGCCCCCACTTAATTTGTAATTTCTAATCTGCTGCTGCTGCTGCTGCT 1680
 DB 1621 TCTCTTGGCCCCCACTTAATTTGTAATTTCTAATCTGCTGCTGCTGCTGCTGCT 1680
 OY 1681 AAATCCAAATATAGTCAACAGTGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
 DB 1681 AAATCCAAATATAGTCAACAGTGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
 OY 1741 ATCTGATTTCTGAGAAACATACAAAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
 DB 1741 ATCTGATTTCTGAGAAACATACAAAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
 OY 1801 TACATTTCCAAAAATATGCTGAGATTTCTGAGTCAATGATGCTGAGAGTCAATTT 1860
 DB 1801 TACATTTCCAAAAATATGCTGAGATTTCTGAGTCAATGATGCTGAGAGTCAATTT 1860
 OY 1861 CACTTGTGGAGCTCAATATGTTTCTGTGACAACTAATCAATGATGCTGCTGCTGCTGCT 1920
 DB 1861 CACTTGTGGAGCTCAATATGTTTCTGTGACAACTAATCAATGATGCTGCTGCTGCTGCT 1920
 OY 1921 AATCAATTCATGAGAAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980

DB 1921 AATCAATTCATGAGAAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
 OY 1981 GATTTGTATTCATTTATCCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040
 DB 1981 GATTTGTATTCATTTATCCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040
 OY 2041 GGATTCATTCATTCAGAGTGTAGTGAAGAGTGTGAGGGAATGGAATGGAAGTGGCC 2100
 DB 2041 GGATTCATTCATTCAGAGTGTAGTGAAGAGTGTGAGGGAATGGAATGGAAGTGGCC 2100
 OY 2101 ACTGTGATGATGCTGCTGAGAGTGTGGAATGAGAGTGTGCTGCTGCTGCTGCTG 2160
 DB 2101 ACTGTGATGATGCTGCTGAGAGTGTGGAATGAGAGTGTGCTGCTGCTGCTGCTG 2160
 OY 2161 GACATCTCAAGCTTTTAAACATTAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2220
 DB 2161 GACATCTCAAGCTTTTAAACATTAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2220
 OY 2221 TTGAATGCAATGAGAGTGTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2280
 DB 2221 TTGAATGCAATGAGAGTGTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2280
 OY 2281 TATTTGGAATGCTGAGTGTAGGAGCCAGCAAGATGTAATATATATATATATATAT 2340
 DB 2281 TATTTGGAATGCTGAGTGTAGGAGCCAGCAAGATGTAATATATATATATATATAT 2340
 RESULT 3
 AF320293 2340 bp mRNA linear PRI 13-DEC-2000
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 2340)
 Berge, K.E., Tian, H., Graf, G.A., Yu, L., Grishin, N.V., Schultz, J.,
 Kwiatkowski, P., Shan, B., Barnes, R., and Hobbs, H.H.
 Accumulation of Dietary Cholesterol in Silesterolemia Caused by
 Mutations in Adjacent ABC Transporters
 Science (2001) In press
 2 (bases 1 to 2340)
 Berge, K.E., Tian, H., Graf, G.A., Yu, L., Grishin, N.V., Schultz, J.,
 Kwiatkowski, P., Shan, B., Barnes, R., and Hobbs, H.H.
 Direct Submission
 Submitted (09-NOV-2000) Molecular Genetics, University of Texas,
 Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd.,
 Dallas, TX 75390-9046, USA
 FEATURES
 source
 1. 2340
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 1. 2340
 /gene="ABCG5"
 /note="ABCG5"
 /note="ATP-binding cassette, subfamily G, member 5"
 /product="ABCG5"
 /protein_id="AAG40003.1"
 /db_xref="GI:11692800"
 /translation="MGDLSSLTLPFGSGNGLVNRGSSSLGAPATAPEPHSLIHNS
 YSVSHRVPMMDLTSCROOQTRQILKQVSLYRSGLIMCIGSSGSKTLLADMSR
 LGRAGTFLGEVYNGRALRREORODPSYVLOSDTLLSTVETLHTLTLIRGN
 PGRORKEVVAEALSLSHADRLIGTSLAGSLSTGERRVSLAQLLDLPKVMLEDE
 PTLGLDQFANQIVLVLVELARNRRLVLTILHQRSELELDFKILILSGELIFGCT
 PAMLEDFNDGCPDPEHNPDPYDLDSVDSQSELELETSRKYOMIESAYKKSNI
 CHRTLNIERMKHLKTLPMVPEFTKDSPGVESKLVGLLRVTRNLVNRKLAIVITRLLO

NLIMGLFLEFVILRVSRNLKGAIDRVLGLYOPFGATPYTGMANVLEPVLRAVSD
 OESODGLXOKOMOMLAVYALHVLPRSVVATMFSSVYCTITLGHPEVAFYSALIA
 PHILGEPLTIVLIGIVONPNVINSVVALISLAVGSGFLNIDEMIPKILISYFT
 FOKTSEILVNERFGLNFTGSSNVSSTTBNMCAFTGIGIOFIEKTCGATSRFTMNF
 LILYSFIPALVILGIVFKIRDLISR*
 BASE COUNT 541 a 601 c 598 g 600 t
 ORIGIN

Query Match 100.0% Score 2340; DB 9; Length 2340;
 Best Local Similarity 100.0% Pred. No. 0;
 Matches 2340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCAGGTGAGCAGCAGGAGGAGTGTGCAAGGGGCTCCCAACTAAGCCACTGTGGGA 60
 DB 1 GTCAGGTGAGCAGCAGGAGGAGTGTGCAAGGGGCTCCCAACTAAGCCACTGTGGGA 60
 QY 61 GGGTCGGCCACAGAAAATTTGCCACGTTTGCTGCTGTGGCCATGGGAGCTCTTC 120
 DB 61 GGGTCGGCCACAGAAAATTTGCCACGTTTGCTGCTGTGGCCATGGGAGCTCTTC 120
 QY 121 ATCTTTGACCCCGGAGGGGTCCATGGGTCTCCAAAGTAAACAGAGGCTCCAGAGCTCCCT 180
 DB 121 ATCTTTGACCCCGGAGGGGTCCATGGGTCTCCAAAGTAAACAGAGGCTCCAGAGCTCCCT 180
 QY 181 GGAGGGGGCTCTGCGCACCGCCCGGAGGCTCAGAGCTGGGACATCTCCATGCTCCTA 240
 DB 181 GGAGGGGGCTCTGCGCACCGCCCGGAGGCTCAGAGCTGGGACATCTCCATGCTCCTA 240
 QY 241 CAGCGTACAGCAGCAGCGCTGAGGCGCTGTGGGACATCATCTTGCGGGCAGAGTGGAC 300
 DB 241 CAGCGTACAGCAGCAGCGCTGAGGCGCTGTGGGACATCATCTTGCGGGCAGAGTGGAC 300
 QY 301 CAGGCAATCCTCAAGATGTCTCTTGTAGTGAAGAGCGGCGAGATGTGCATCT 360
 DB 301 CAGGCAATCCTCAAGATGTCTCTTGTAGTGAAGAGCGGCGAGATGTGCATCT 360
 QY 361 AGGAAGCTCAGGCTCCGGGAAAACACAGCTGTGAGCCATGTCGGGAGGCTGGGGCG 420
 DB 361 AGGAAGCTCAGGCTCCGGGAAAACACAGCTGTGAGCCATGTCGGGAGGCTGGGGCG 420
 QY 421 CGGCGGAGCTCTCTGCGGAGGAGTGTAGTGAACGGCGGGCGTGGCGCGGAGCACTT 480
 DB 421 CGGCGGAGCTCTCTGCGGAGGAGTGTAGTGAACGGCGGGCGTGGCGCGGAGCACTT 480
 QY 481 CCAAGACTGCTTCTCTACGTCCTGAGAGCGAACCCCTGTGAGCAAGCTCACCCTGCG 540
 DB 481 CCAAGACTGCTTCTCTACGTCCTGAGAGCGAACCCCTGTGAGCAAGCTCACCCTGCG 540
 QY 541 CGAGAGCTGCTACACCGCGCTGCGGCAATCCGCGGCAATCCGAGCTCTTCCA 600
 DB 541 CGAGAGCTGCTACACCGCGCTGCGGCAATCCGCGGCAATCCGAGCTCTTCCA 600
 QY 601 GAAGAAGTGGAGCGGCTATGGCAGAGTGAAGTGAAGCCATGTGGCAACAGCGACTGT 660
 DB 601 GAAGAAGTGGAGCGGCTATGGCAGAGTGAAGTGAAGCCATGTGGCAACAGCGACTGT 660
 QY 661 TGGCAACTACAGCTTGGGGGCAATTCACAGGGTGAAGCGCGCGGGTCTCCATCCGAGC 720
 DB 661 TGGCAACTACAGCTTGGGGGCAATTCACAGGGTGAAGCGCGCGGGTCTCCATCCGAGC 720
 QY 721 CGAGCTGCTCCAGAGATCTTAAGTCAATGCTGTTTGAATGAGCAACACAGGCTTGACTG 780
 DB 721 CGAGCTGCTCCAGAGATCTTAAGTCAATGCTGTTTGAATGAGCAACACAGGCTTGACTG 780
 QY 781 CATGACTGCTAATCATATGTGTGCTCTGAGTGAAGTGAAGCGCTGAGAGCAAGCAATGT 840
 DB 781 CATGACTGCTAATCATATGTGTGCTCTGAGTGAAGTGAAGCGCTGAGAGCAAGCAATGT 840
 QY 841 GGTTCACACATTCACAGCCCGTCTGAGCTTTTTCAGCTCTTTGACAAAATTTGCCAT 900
 DB 841 GGTTCACACATTCACAGCCCGTCTGAGCTTTTTCAGCTCTTTGACAAAATTTGCCAT 900
 QY 901 CCTGAGCTTGGAGAGCTGATTTTCTGTGGACGCCAGCGGAATGCTGATTTCTTCAA 960

DB 901 CCTGAGCTTGGAGAGCTGATTTTCTGTGGACGCCAGCGGAATGCTGATTTCTTCAA 960
 QY 961 TGACGCGGCTTACCTTGTCTCTGAACATTCAAACCTTTTGACTTCTATAGGAGCTGAC 1020
 DB 961 TGACGCGGCTTACCTTGTCTCTGAACATTCAAACCTTTTGACTTCTATAGGAGCTGAC 1020
 QY 1021 GTCAGTGAATACCCCAAGCAGGAGGGAATTAAGAACTCCAGAGAGTCAATGAT 1080
 DB 1021 GTCAGTGAATACCCCAAGCAGGAGGGAATTAAGAACTCCAGAGAGTCAATGAT 1080
 QY 1081 AGAATCTGCTACAAAGAAATGAGCAATTTGTATATAAATTTGAAGATATTAAGAAAT 1140
 DB 1081 AGAATCTGCTACAAAGAAATGAGCAATTTGTATATAAATTTGAAGATATTAAGAAAT 1140
 QY 1141 GAAACACCTGAAACGTTTACCAATGTTCTTTCAAACCAAAATTTCTCTGAGTTT 1200
 DB 1141 GAAACACCTGAAACGTTTACCAATGTTCTTTCAAACCAAAATTTCTCTGAGTTT 1200
 QY 1201 CTCTAACTGGGTGTTCTCTGAGAGGTGACAAAGAACTGTGAGAAATTAAGTGGC 1260
 DB 1201 CTCTAACTGGGTGTTCTCTGAGAGGTGACAAAGAACTGTGAGAAATTAAGTGGC 1260
 QY 1261 AGTATTAACGGCTCTCTTCTAGAAATCTGATGAGTTTGTCTCTTTCTTCTGTTCT 1320
 DB 1261 AGTATTAACGGCTCTCTTCTAGAAATCTGATGAGTTTGTCTCTTTCTTCTGTTCT 1320
 QY 1321 GCGGCTCGAAGCAATGTCTAAAGGTGCTATACAGAGCCGCTGAGTCTCTTTACCA 1380
 DB 1321 GCGGCTCGAAGCAATGTCTAAAGGTGCTATACAGAGCCGCTGAGTCTCTTTACCA 1380
 QY 1381 GTTTGTGGGCGCACCCGCTACAGAGCATCTGAAGCTGAATCTGTTCCCGTCT 1440
 DB 1381 GTTTGTGGGCGCACCCGCTACAGAGCATCTGAAGCTGAATCTGTTCCCGTCT 1440
 QY 1441 GCGAGCTGTCAGCAGCAGAGAGTCAAGAGCGGCTCTACAGAAAGTGGCAGATAGCT 1500
 DB 1441 GCGAGCTGTCAGCAGCAGAGAGTCAAGAGCGGCTCTACAGAAAGTGGCAGATAGCT 1500
 QY 1501 GCGCTATGACATGACGCTCTCCCTCAGAGGTTGTGCCACATGATTTTACAGAGGT 1560
 DB 1501 GCGCTATGACATGACGCTCTCCCTCAGAGGTTGTGCCACATGATTTTACAGAGGT 1560
 QY 1561 GTGCTACTGAGAGCTGGGCTTACATCTGAGGTTGCCAGATGATTTTCTGCTGC 1620
 DB 1561 GTGCTACTGAGAGCTGGGCTTACATCTGAGGTTGCCAGATGATTTTCTGCTGC 1620
 QY 1621 TCTCTGGCCCCACATTAATTTGTTGAATTTCTAACCTTGTGCTACTTGGTATGCTCA 1680
 DB 1621 TCTCTGGCCCCACATTAATTTGTTGAATTTCTAACCTTGTGCTACTTGGTATGCTCA 1680
 QY 1681 AAATCCAAATATATAGTCAACAGTGTAGTGGCTGCTGCTCATTTGCGGGGCTGTTGG 1740
 DB 1681 AAATCCAAATATATAGTCAACAGTGTAGTGGCTGCTGCTCATTTGCGGGGCTGTTGG 1740
 QY 1741 ATCTGATTTCTCAGAAACATACAGAAATGCCATTTCTTTAAATCATAGTATTT 1800
 DB 1741 ATCTGATTTCTCAGAAACATACAGAAATGCCATTTCTTTAAATCATAGTATTT 1800
 QY 1801 TACATTCCAAAATATATGAGTGAATTTCTTGTAGTCAATGAGTCTGAGGAGCTGATTT 1860
 DB 1801 TACATTCCAAAATATATGAGTGAATTTCTTGTAGTCAATGAGTCTGAGGAGCTGATTT 1860
 QY 1861 CACTTGTGAGCTCAAAATGTTCTGTGACAACTAATTCAAATGTGCTTCACTCAAG 1920
 DB 1861 CACTTGTGAGCTCAAAATGTTCTGTGACAACTAATTCAAATGTGCTTCACTCAAG 1920
 QY 1921 AATTCAATCATTTGAGAAAAACCTGCGAGGTGCAATCTAGATTCACAAATGAACTTCT 1980
 DB 1921 AATTCAATCATTTGAGAAAAACCTGCGAGGTGCAATCTAGATTCACAAATGAACTTCT 1980
 QY 1981 GATTTGTATTCATTTATTCAGCTCTGTCACATCAGAAATGTTGTTTCAAATAG 2040

Db	1981	GATTTCGATTCATTATTATTCACGCTCTTGTCATCCCTAGGAATAGTGTGTTTCAAAATGAG	2040
Oy	2041	GGATCATCTCATTTACAGGTAGTGAAGCCATGGCTGGGAAATGGAGTGAAGTGGCG	2100
Db	2041	GGATCATCTCATTTACAGGTAGTGAAGCCATGGCTGGGAAATGGAGTGAAGTGGCG	2100
Oy	2101	ACTGTGCATGACTGCTCTGAACGCTGTGAATAGAGTGCATGTATTTCTTCTTGGACAG	2160
Db	2101	ACTGTGCATGACTGCTCTGAACGCTGTGAATAGAGTGCATGTATTTCTTCTTGGACAG	2160
Oy	2161	GACATCTCAAGCTTTTATACCATTTAGACTCCATTTGTGCTCTTGATCCACAGGCC	2220
Db	2161	GACATCTCAAGCTTTTATACCATTTAGACTCCATTTGTGCTCTTGATCCACAGGCC	2220
Oy	2221	TTGATGCAATGGAAGTGGTTATATAGTCCCTGCTCTTACAACTTGCAGGACATGTGCT	2280
Db	2221	TTGATGCAATGGAAGTGGTTATATAGTCCCTGCTCTTACAACTTGCAGGACATGTGCT	2280
Oy	2281	TATTTGGAAATTTGACTGACGACGCGACCCAGATGTAAATATTCATTAAGCTATGGG	2340
Db	2281	TATTTGGAAATTTGACTGACGACGCGACCCAGATGTAAATATTCATTAAGCTATGGG	2340

RESULT 4
AF312715

LOCUS
DEFINITION Homo sapiens sterolin (ABCG5) mRNA, complete cds.

ACCESSION
AF312715.2 GI:14423628

VERSION
AF312715.2

KEYWORDS
Homo sapiens (human)

SOURCE
ORGANISM Homo sapiens

REFERENCE
1. (bases 1 to 2740)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

AUTHORS
Lee,M.H., Lu,K., Hazard,S., Yu,H., Shulien,S., Hidaka,H.,
Kojima,H., Allikmets,R., Sakuma,N., Pegoraro,R., Srivastava,A.K.,
Salem,G., Dean,M. and Patel,S.B.
Identification of a gene, ABCG5, important in the regulation of
dietary cholesterol absorption
Nat. Genet. 27 (1), 79-83 (2001)

TITLE
JOURNAL
MEDLINE
PUBMED
1138003

REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases: 1 to 2740)
Lu,K., Lee,M.-H. and Patel,S.B.
Direct Submission
Submitted (12-OCT-2000) Division of Endocrinology, Diabetes and
Medical Genetics, Medical University of South Carolina, 114 Doughty
St., STB541, Charleston, SC 29403, USA
On Jun 14, 2001 this sequence version replaced gi:12382303.
Location/Qualifiers
1..2740

2740 bp mRNA linear PRI 14-JUN-2001

NMIGLFLLEFLFVRNSVNLKGAITODPVGLLYXQVCGATPPTGMLNANVNLPEVLRAVSD
 QESODDGLYOKQOMMLAVALHWLPFSVVALNLFSSVCITWGLHBNARFEGFSALLA
 PHILGEITLVLCTIVONPNVINSVALLSIAGVAGSGFLRNQIENHPIPKIISYPTT
 FQKSEFLVNIENYEGLTFTGSSNSVSTTNPMCAFTQGIQFIKTCPGATSRFTNMF
 LILSFELPALVILGIVYFKIRHDIISR
 BASE COUNT 669 a 670 c 702 g 699 t
 ORIGIN

Query Match	100.0%;	Score 2340;	DB 9;	Length 2740;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2340;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

OY	1	IGCAGAGTGGAGGAGGAGGGGACGTGTGGCACAGGGGCTCCCAACTBACACCACTGTGGGGA	60
Db	35	GTCAAGTGGAGGAGGAGGAGGTGTGGCACAGGGGCTCCCAACTBACACCACTGTGGGGA	94
OY	61	GGGTCCGGCCACACAGAAATTTTGCCACGTTTGTGCTGTGGGCATGGGTGACCTTTC	120
Db	95	GGGTCCGGCCACACAGAAATTTTGCCACGTTTGTGCTGTGGGCATGGGTGACCTTTC	154
OY	121	ATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAACTAAACAGAGGCTCCAGAGTCCCT	180
Db	155	ATCTTTGACCCCCGGAGGGTCCATGGGTCTCCAACTAAACAGAGGCTCCAGAGTCCCT	214
OY	181	GGAGGGGGGTCTCTGGCACCCCGCCGGAGACCTTCACAGCCTGGGCATCTTCATGCTCTTA	240
Db	215	GGAGGGGGGTCTCTGGCACCCCGCCGGAGACCTTCACAGCCTGGGCATCTTCATGCTCTTA	274
OY	241	CAGCTTCACGCCACCCGCTGAGGGCCCTGGTGGGACATCAATCTTGGCGGGACAGTGGAC	300
Db	275	CAGCTTCACGCCACCCGCTGAGGGCCCTGGTGGGACATCAATCTTGGCGGGACAGTGGAC	334
OY	301	CAGGAGATTCCTCAAAAGATGTCTCTTGTACGTGGAGAGCGGGACAGATCATGTTCATCT	360
Db	335	CAGGAGATTCCTCAAAAGATGTCTCTTGTACGTGGAGAGCGGGACAGATCATGTTCATCT	394
OY	361	AGGAAGCTCAGGCTCCGGGAAAACACAGCCTCTGTGAGGCCATGTCCGGGAGCTGGGGCG	420
Db	395	AGGAAGCTCAGGCTCCGGGAAAACACAGCCTCTGTGAGGCCATGTCCGGGAGCTGGGGCG	454
OY	421	CGCGGGGACCTTCTCGGGGAGGTGTATGTAAAGGCGGGCGCTGGCGGGGAGCAAGTT	480
Db	455	CGCGGGGACCTTCTCGGGGAGGTGTATGTAAAGGCGGGCGCTGGCGGGGAGCAAGTT	514
OY	481	CCAGGACTGCTTCTCTTACGTCGTGCAGAGAGGACACCCTGCTGAGACGCTTCACCTGTGC	540
Db	515	CCAGGACTGCTTCTCTTACGTCGTGCAGAGAGGACACCCTGCTGAGACGCTTCACCTGTGC	574
OY	541	CGAAGCGTGCACATCACCCGCGTGTGCGCATCCCGCGGGCAATCCCGGCTCTTCCA	600
Db	575	CGAAGCGTGCACATCACCCGCGTGTGCGCATCCCGCGGGCAATCCCGGCTCTTCCA	634
OY	601	GAAAGAGGTGAGGCGCGTCAATGGGACAGCTGAGCTBAGTCATGTGGCACAACCGACTGAT	660
Db	635	GAAAGAGGTGAGGCGCGTCAATGGGACAGCTGAGCTBAGTCATGTGGCACAACCGACTGAT	694
OY	661	TGGCACTACAGCTTTGGGGGGCATTTTCCAGGGGTGAGCGGCGCGGGGTCCCATGCGAGC	720
Db	695	TGGCACTACAGCTTTGGGGGGCATTTTCCAGGGGTGAGCGGCGCGGGGTCTCCATGCGAGC	754
OY	721	CCAGCTGCTCCAGGATCCTAAGGTCAAGTCATGATGCTTTGATGAGCCAAACAGAGCCTGAGCTG	780
Db	755	CCAGCTGCTCCAGGATCCTAAGGTCAAGTCATGATGCTTTGATGAGCCAAACAGAGCCTGAGCTG	814
OY	781	CATACAGCTAATAGATTTGCTCCCTCTGCTGGTGGAACTGGCTTCGACAGAACCGAATTTGT	840
Db	815	CATACAGCTAATAGATTTGCTCCCTCTGCTGGTGGAACTGGCTTCGACAGAACCGAATTTGT	874
OY	841	GGTTCACCAATTCACAGCCCGTGTCTGAGCTTTTTCAGCTTTTGAACAAATTTGCCAT	900
Db	875	GGTTCACCAATTCACAGCCCGTGTCTGAGCTTTTTCAGCTTTTGAACAAATTTGCCAT	934
OY	901	CTGAGACTTGGAGAGCTGATTTTCTGTGGCACGCCAGCGGAATCTTGATTTCTTCAA	960

Db	935	CTGAGCTTGGAGAGCTGATTTTCTGTGACGCCGCGGAAATGCTGATTTCTTCA	994
Qy	961	TGACTGGGTTTACCTTGTCTGTACATCAAAACCTTTTACTTCTATATGACCTGAC	1020
Db	995	TGACTGGGTTTACCTTGTCTGTACATCAAAACCTTTTACTTCTATATGACCTGAC	1054
Qy	1021	GTCAGTGGATACCCAAAGCAAGAACGGAAATAGAAACCTCCAGAGAGTCCAGATGAT	1080
Db	1055	GTCAGTGGATACCCAAAGCAAGAACGGAAATAGAAACCTCCAGAGAGTCCAGATGAT	1114
Qy	1081	AGATCTGCTTACAGAAATACAGCAATTTGTCAATAAACTTTGAAGAAATTTGAAGAAAT	1140
Db	1115	AGATCTGCTTACAGAAATACAGCAATTTGTCAATAAACTTTGAAGAAATTTGAAGAAAT	1174
Qy	1141	GAAGACCTGAAACGTTTACCAATGTTCTTCAAAACCAAGATTTCTCTGAGTTT	1200
Db	1175	GAAGACCTGAAACGTTTACCAATGTTCTTCAAAACCAAGATTTCTCTGAGTTT	1234
Qy	1201	CTCTAAACTGGGTCTCTCTGAGAGAGTACAGAACTGTGTGAGAAATAGCTGGC	1260
Db	1235	CTCTAAACTGGGTCTCTCTGAGAGAGTACAGAACTGTGTGAGAAATAGCTGGC	1294
Qy	1261	AGTATTTACGCTCTCTCTGAGAAATGATGATGGTTTGTCTCTCTCTCTCTCTCTCTCT	1320
Db	1295	AGTATTTACGCTCTCTCTCTGAGAAATGATGATGGTTTGTCTCTCTCTCTCTCTCTCTCT	1354
Qy	1321	GGGGGTCCGAAGAATGTCTTAAGGTGTATCCAGAGACCGGTGAGTCTCTTTACCA	1380
Db	1355	GGGGGTCCGAAGAATGTCTTAAGGTGTATCCAGAGACCGGTGAGTCTCTTTACCA	1414
Qy	1381	GTTTGTGGGGCCACCCCGTACACAGCATGTGAACGGCTGTAATCTGTTTCCCGTGT	1440
Db	1415	GTTTGTGGGGCCACCCCGTACACAGCATGTGAACGGCTGTAATCTGTTTCCCGTGT	1474
Qy	1441	GCGAGCTGTACGAGCAGCAGAGAGTACAGAGCGGCTCTACCAAGTGGCAGATGATGT	1500
Db	1475	GCGAGCTGTACGAGCAGCAGAGAGTACAGAGCGGCTCTACCAAGTGGCAGATGATGT	1534
Qy	1501	GGCCTATGACCTGACGCT	1560
Db	1535	GGCCTATGACCTGACGCT	1594
Qy	1561	GTCCTACTGAGACCTGGGCTTACATCTGAGTGGCCGATTTGATATTTTCTCTCTCT	1620
Db	1595	GTCCTACTGAGACCTGGGCTTACATCTGAGTGGCCGATTTGATATTTTCTCTCTCT	1654
Qy	1621	TCTCTTGGCCCCCCT	1680
Db	1655	TCTCTTGGCCCCCCT	1714
Qy	1681	AAATCCAAATATAGTCAACAGTGTAGTGTCTGTCTCTCTCTCTCTCTCTCTCTCTCT	1740
Db	1715	AAATCCAAATATAGTCAACAGTGTAGTGTCTGTCTCTCTCTCTCTCTCTCTCTCTCT	1774
Qy	1741	ATCTGATTTCTCTAGAAACATACAAAGAAATGCCANTCTCTTTAAATATACATTTT	1800
Db	1775	ATCTGATTTCTCTAGAAACATACAAAGAAATGCCANTCTCTTTAAATATACATTTT	1834
Qy	1801	TACATTTCCAAAATATATGAGTGTAGTGTCTGTAGTGTCTGTAGTGTCTGTAGTGTCT	1860
Db	1835	TACATTTCCAAAATATATGAGTGTAGTGTCTGTAGTGTCTGTAGTGTCTGTAGTGTCT	1894
Qy	1861	CACCTTGGCAGCTCAAAATGTTTCTGTGCAACTAATCCAAATGTCCTTCTCTCTCTCT	1920
Db	1895	CACCTTGGCAGCTCAAAATGTTTCTGTGCAACTAATCCAAATGTCCTTCTCTCTCTCT	1954
Qy	1921	AATTCATTTCAATGAGAAACCTGGCCAGGTGCAACATAGATTTCAACAATATCTTCT	1980
Db	1955	AATTCATTTCAATGAGAAACCTGGCCAGGTGCAACATAGATTTCAACAATATCTTCT	2014
Qy	1981	GATTTTGTATTTATTTATTTCTGAGCTCTTGTCTATCTCTAGAAATAGTTGTTTCAAAATAG	2040
Db	2015	GATTTTGTATTTATTTATTTCTGAGCTCTTGTCTATCTCTAGAAATAGTTGTTTCAAAATAG	2074
Qy	2041	GGATCATCTTATTTAGCAGGTAGTGAAGCCATGCTGGGAAATGGAAGTGAAGCTGCC	2100
Db	2075	GGATCATCTTATTTAGCAGGTAGTGAAGCCATGCTGGGAAATGGAAGTGAAGCTGCC	2134
Qy	2101	ACTGTGATGACAGCTCTGTGAAGCTGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT	2160
Db	2135	ACTGTGATGACAGCTCTGTGAAGCTGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT	2194
Qy	2161	GACATCTCAAGCTCTTTTAAACATTAAGTCACTATTTGTGCTCTTGGATCCAGAGGCC	2220
Db	2195	GACATCTCAAGCTCTTTTAAACATTAAGTCACTATTTGTGCTCTTGGATCCAGAGGCC	2254
Qy	2221	TTGAATGCAATGGAAGTGTATATAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2280
Db	2255	TTGAATGCAATGGAAGTGTATATAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2314
Qy	2281	TATTTGGAATTTGACTGAGCGGACCCAGCAAGATGTAATATATTTCAATTAACCTATGG	2340
Db	2315	TATTTGGAATTTGACTGAGCGGACCCAGCAAGATGTAATATATTTCAATTAACCTATGG	2374
RESULT	5		
AX456520			
LOCUS	AX456520	2516 bp	DNA linear
DEFINITION	Sequence 42 from Patent WO0227016.		
ACCESSION	AX456520		
VERSION	AX456520.1	GI:21715410	
KEYWORDS			
SOURCE			
ORGANISM	synthetic construct		
REFERENCE	1		
AUTHORS	Patel, S.B. and Dean, M.		
TITLE	Gene involved in dietary sterol absorption and excretion and uses		
JOURNAL	therefor		
FEATURES	Patent: WO 0227016-A 42 04-APR-2002;		
source	THE DEPARTMENT OF HEALTH AND HUMAN SERVICES (US) ; Patel,		
	Shallendra B. (US) ; Dean, Michael (US)		
	Location/Qualifiers		
	1..2516		
	/organism="synthetic construct"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:32630"		
	/note="Primer"		
BASE COUNT	601 a 631 c 636 g 648 t		
ORIGIN			
Query Match	99.98; Score 2338.4; DB 6; Length 2516;		
Best Local Similarity	100.0%; Pred. No. 0;		
Matches 2339; Conservative	0; Mismatches 1; Indels 0; Gaps 0;		
Qy	1	GTCAGTGGAGAGAGGAGGAGCTGTGCGACAGGAGCTCCCACTGAAGCACTGTGGGA	60
Db	35	GTCAGTGGAGAGAGGAGGAGCTGTGCGACAGGAGCTCCCACTGAAGCACTGTGGGA	94
Qy	61	GGTCCGCGACACAGAAATTTGGCCAGCTTTGTGCTGCTGTTGGCCATGGGTGACCTTC	120
Db	95	GGTCCGCGACACAGAAATTTGGCCAGCTTTGTGCTGCTGTTGGCCATGGGTGACCTTC	154
Qy	121	ATCTTTGACCCCGGAGGCTCATGGGTCTCCAACTAAACAGAGGCTCCAGAGCTCCCT	180
Db	155	ATCTTTGACCCCGGAGGCTCATGGGTCTCCAACTAAACAGAGGCTCCAGAGCTCCCT	214
Qy	181	GGAGGGGCTCTCTGACAGCGCCCGGAGGCTCAAGCTGGGCAATCTTCATGCTCTCA	240
Db	215	GGAGGGGCTCTCTGACAGCGCCCGGAGGCTCAAGCTGGGCAATCTTCATGCTCTCA	274
Qy	241	CAGCTGACGACCGGCTGAGGCTGTGGGAGATCAATCTTGGCGGAGCAGCTGGAC	300
Db	275	CAGCTGACGACCGGCTGAGGCTGTGGGAGATCAATCTTGGCGGAGCAGCTGGAC	334

OY 301 CAGCAGATCTCTAAAGATGTCTCTTGTACGTGAGAGCGGCGAGATCATGTGCATCTT 360
|||||
DB 335 CAGGCGAGATCTCTAAAGATGTCTCTTGTACGTGAGAGCGGCGAGATCATGTGCATCTT 394
|||||
OY 361 AGGAAGCTCAGGCTCCGGGAAAACACGCGTGTGAGAGCGGCGAGATCATGTGCATCTT 420
|||||
DB 395 AGGAAGCTCAGGCTCCGGGAAAACACGCGTGTGAGAGCGGCGAGATCATGTGCATCTT 454
|||||
OY 421 CCGGGGAGACCTTCTCGGGGAGGTGTATGTGAACGGCCGGGCGCTCGCGGGAGCAGTT 480
|||||
DB 455 CCGGGGAGACCTTCTCGGGGAGGTGTATGTGAACGGCCGGGCGCTCGCGGGAGCAGTT 514
|||||
OY 481 CCGAGACTGCTTCTCTAGCTCTGTGAGAGCGAGACCGCTGTGAGAGCGCTCACCGTGGC 540
|||||
DB 515 CCGAGACTGCTTCTCTAGCTCTGTGAGAGCGAGACCGCTGTGAGAGCGCTCACCGTGGC 574
|||||
OY 541 CCGAGACTGCTCTACACCGCGCTGTGAGAGCGAGACCGCGGCGGCGGCGGCGCTCTTCCA 600
|||||
DB 575 CCGAGACTGCTCTACACCGCGCTGTGAGAGCGAGACCGCGGCGGCGGCGGCGCTCTTCCA 634
|||||
OY 601 GAAGAAGTGAAGGCGGCTGATGAGAGCTGAGTGTGAGCATGTGAGAGCGGCGGCGCTGAT 660
|||||
DB 635 GAAGAAGTGAAGGCGGCTGATGAGAGCTGAGTGTGAGCATGTGAGAGCGGCGGCTGAT 694
|||||
OY 661 TGGCAACTACAGCTTGGGGGGCATTTTCCACGGGTGAGCGCGCGGGTCTCCATCCGACG 720
|||||
DB 695 TGGCAACTACAGCTTGGGGGGCATTTTCCACGGGTGAGCGCGCGGGTCTCCATCCGACG 754
|||||
OY 721 CCGAGCTGCGAGAGATCTTAAGGTGATGTGTTGATGAGCCACACAGCGCTGACG 780
|||||
DB 755 CCGAGCTGCGAGAGATCTTAAGGTGATGTGTTGATGAGCCACACAGCGCTGACG 814
|||||
OY 781 CATGACTGCTTAATCAGATTTGTGCTCTCTGTGAGAGCGGCTGAGAGACCGGATTTGT 840
|||||
DB 815 CATGACTGCTTAATCAGATTTGTGCTCTCTGTGAGAGCGGCTGAGAGACCGGATTTGT 874
|||||
OY 841 GGTTCCTACCAATTCACAGCGCGGCTGTGAGCTTTTTCAGCTTTTTCAGAAAATTTCCAT 900
|||||
DB 875 GGTTCCTACCAATTCACAGCGCGGCTGTGAGCTTTTTCAGCTTTTTCAGAAAATTTCCAT 934
|||||
OY 901 CCGAGCTGCGAGAGATCTTAAGGTGATGTGTTGATGAGCCACACAGCGCTGACG 960
|||||
DB 935 CCGAGCTGCGAGAGATCTTAAGGTGATGTGTTGATGAGCCACACAGCGCTGACG 994
|||||
OY 961 TGAAGTGGGTTACCTTGTCTGTAACATTTCAAAACCTTTTGAATTTATATGAGACTGAC 1020
|||||
DB 995 TGAAGTGGGTTACCTTGTCTGTAACATTTCAAAACCTTTTGAATTTATATGAGACTGAC 1054
|||||
OY 1021 GTCAAGTGAATCCCAAGAGAGGAAATAGAAATCTCCAGAGAGTCCAGATGAT 1080
|||||
DB 1055 GTCAAGTGAATCCCAAGAGAGGAAATAGAAATCTCCAGAGAGTCCAGATGAT 1114
|||||
OY 1081 AGAATCTGCTCAAGAAATCAGCAATTTGTATATAAATCTTGAAGATATTTGAAGAAT 1140
|||||
DB 1115 AGAATCTGCTCAAGAAATCAGCAATTTGTATATAAATCTTGAAGATATTTGAAGAAT 1174
|||||
OY 1141 GAAGAAGCTGAAGAGTTTACCAATGTTCTTTCAAAACCAAGATTTCTCGAGTTT 1200
|||||
DB 1175 GAAGAAGCTGAAGAGTTTACCAATGTTCTTTCAAAACCAAGATTTCTCGAGTTT 1234
|||||
OY 1201 CTCTAACTGGGTGTTCTCTGAGAGAGTGAACAAGAACTTGTGAGAAATTAAGCTGAC 1260
|||||
DB 1235 CTCTAACTGGGTGTTCTCTGAGAGAGTGAACAAGAACTTGTGAGAAATTAAGCTGAC 1294
|||||
OY 1261 AGTGAATTAAGGCTCTCAATCTGATGATGATGTTGTTGCTCTCTCTCTCTCTCTCTCT 1320
|||||
DB 1295 AGTGAATTAAGGCTCTCTCAATCTGATGATGATGTTGTTGCTCTCTCTCTCTCTCTCTCT 1354
|||||
OY 1321 GCGGGTCCGAAGCAATGTCTTAAGAGTCTATCCAGAGACCGGATAGTCTCTTTACCA 1380
|||||
DB 1355 GCGGGTCCGAAGCAATGTCTTAAGAGTCTATCCAGAGACCGGATAGTCTCTTTACCA 1414
|||||
OY 1381 GTTTGTGGGGCCACCCCGTACACAGGAGCTGTAACGCTGTGAATCTGTCTCTCTCTCTCT 1440
|||||

DB 1415 GTTTGTGGGGCCACCCCGTACACAGGCAATGCTGAACGCTGTGAATCTTTCCCTGCT 1474
|||||
OY 1441 GCGAGCTGTACAGGACACAGAGAGTGAAGAGCGGCTCTACACAGAACTGCGACATGTGCT 1500
|||||
DB 1475 GCGAGCTGTACAGGACACAGAGAGTGAAGAGCGGCTCTACACAGAACTGCGACATGTGCT 1534
|||||
OY 1501 GCGCTATGCACTGACAGTCTCTCCCTTCAAGCGGTTGTGACACATATTTTACAGAGTGT 1560
|||||
DB 1535 GCGCTATGCACTGACAGTCTCTCCCTTCAAGCGGTTGTGACACATATTTTACAGAGTGT 1594
|||||
OY 1561 GTGCTACTGAGAGCGTGGGCTTACATCTGAGTTGCGCCGATTTGATATTTTCTCTCTG 1620
|||||
DB 1595 GTGCTACTGAGAGCGTGGGCTTACATCTGAGTTGCGCCGATTTGATATTTTCTCTCTG 1654
|||||
OY 1621 TCTCTTGGCCCCCACTTAATTGTGAATTTCTACTCTTGTGCTACTTGTATGCTGCA 1680
|||||
DB 1655 TCTCTTGGCCCCCACTTAATTGTGAATTTCTACTCTTGTGCTACTTGTATGCTGCA 1714
|||||
OY 1681 AATTCGAATTAATGATCAAGAGTGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
|||||
DB 1715 AATTCGAATTAATGATCAAGAGTGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1774
|||||
OY 1741 ATCTGATTCCTCAGAAACATACAAAGAAATGCCATTTCTTTTAAATCATCAGTATTT 1800
|||||
DB 1775 ATCTGATTCCTCAGAAACATACAAAGAAATGCCATTTCTTTTAAATCATCAGTATTT 1834
|||||
OY 1801 TACATTCAGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1860
|||||
DB 1835 TACATTCAGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1894
|||||
OY 1861 CACTGTGGCAGCTCAAAATGTTTGTGTGCAACTATCCAAATGAGTGTGCTTACGACTGA 1920
|||||
DB 1895 CACTGTGGCAGCTCAAAATGTTTGTGTGCAACTATCCAAATGAGTGTGCTTACGACTGA 1954
|||||
OY 1921 AATTCATTCATTAAGAAACCTGCGCAGAGTCAACATATGATTAATCAATGATTTCT 1980
|||||
DB 1955 AATTCATTCATTAAGAAACCTGCGCAGAGTCAACATATGATTAATCAATGATTTCT 2014
|||||
OY 1981 GATTTTGTATTCATTTATTCAGCTCTGTGTCATCTTGAAGATAGTTGTTTCAAAATAG 2040
|||||
DB 2015 GATTTTGTATTCATTTATTCAGCTCTGTGTCATCTTGAAGATAGTTGTTTCAAAATAG 2074
|||||
OY 2041 GATTCATTCATTAAGAGTGTGAAGAGCAATGCGTGGGAAATGGAATGAACTGCGG 2100
|||||
DB 2075 GATTCATTCATTAAGAGTGTGAAGAGCAATGCGTGGGAAATGGAATGAACTGCGG 2134
|||||
OY 2101 ACTGTGATGATGCTGTGTAAGTCTGAAGATGAGAGTGTGATGATGATGATGATGATG 2160
|||||
DB 2135 ACTGTGATGATGCTGTGTAAGTCTGAAGATGAGAGTGTGATGATGATGATGATGATG 2194
|||||
OY 2161 GATTCATTCATTTTAAACATTAAGACTCCATTTGTGCTTGTGATCAAGAGGCG 2220
|||||
DB 2195 GATTCATTCATTTTAAACATTAAGACTCCATTTGTGCTTGTGATCAAGAGGCG 2254
|||||
OY 2221 TTGAATGCAATGGAAGTGGTTTATGATGCTGCTGCTTCAACATCTGAGAGGCAATG 2280
|||||
DB 2255 TTGAATGCAATGGAAGTGGTTTATGATGCTGCTTCAACATCTGAGAGGCAATG 2314
|||||
OY 2281 TATTTGGAATTTGACTGAGGAGGACCAAGATTAATTAATTAATTAATTAATTAATTAAT 2340
|||||
DB 2315 TATTTGGAATTTGACTGAGGAGGACCAAGATTAATTAATTAATTAATTAATTAATTAAT 2374
|||||

RESULT 6
AX456519 1920 bp DNA linear PAR 06-JUL-2002
LOCUS AX456519
DEFINITION Sequence 41 from Patent WO0227016.
ACCESSION AX456519
VERSION AX456519.1 GI:21715409
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct

REFERENCE 1 artificial sequences.
 AUTHORS Patel,S.B. and Dean,M.
 TITLE Gene involved in dietary sterol absorption and excretion and uses therefor
 JOURNAL Patent: WO 0227016-A 41 04-APR-2002;
 THE DEPARTMENT OF HEALTH AND HUMAN SERVICES (US) ; Patel, Shalendra B. (US) ; Dean, Michael (US)
 FEATURES Location/Qualifiers
 source 1..1920
 /organism="synthetic construct"
 /mol_type="genomic DNA"
 /db_xref="taxon:32630"
 /note="Primer"

BASE COUNT 440 a 503 c 486 g 491 t
 ORIGIN

Query Match 82.1% Score 1920: DB 6: Length 1920:
 Best Local Similarity 100.0%: Prid. No. 0:
 Matches 1920: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 143 ATGGGCTCCAAAGTAACAGAGGCTCCAGAGCTCCGAGAGGGGGCTCCGACCGCC 202
 DB 1 ATGGGCTCCAAAGTAACAGAGGCTCCAGAGCTCCGAGAGGGGGCTCCGACCGCC 60
 OY 203 CCGAGGCTTACAGAGCTGGGCAATCCCTCATGCTCTTACAGAGCTTACAGAGCTGAG 262
 DB 61 CCGAGGCTTACAGAGCTGGGCAATCCCTCATGCTCTTACAGAGCTTACAGAGCTGAG 120
 OY 263 CCGTGGGGAGACATCATCTTGGCCGACAGTGGAGCCAGAGCTTCAAGATATGTC 322
 DB 121 CCGTGGGGAGACATCATCTTGGCCGACAGTGGAGCCAGAGCTTCAAGATATGTC 180
 OY 323 TCCCTTACGTGAGAGCCGGGAGATCATGTGCATCTTACAGAGCTTACAGAGCTTCCGGGAAA 382
 DB 181 TCCCTTACGTGAGAGCCGGGAGATCATGTGCATCTTACAGAGCTTACAGAGCTTCCGGGAAA 240
 OY 383 ACCAGGCTGCTGAGAGCCCATATGTCGGGAGAGCTGGGGGCGCGGAGCTTCTGGGGAG 442
 DB 241 ACCAGGCTGCTGAGAGCCCATATGTCGGGAGAGCTGGGGGCGCGGAGCTTCTGGGGAG 300
 OY 443 GTGTATGTGAAGCGCGGGGCGCTGCGCCCGGAGACAGTTCAGAGCTGCTTCTTACGTC 502
 DB 301 GTGTATGTGAAGCGCGGGGCGCTGCGCCCGGAGACAGTTCAGAGCTGCTTCTTACGTC 360
 OY 503 CTGCAAGAGGACACCTGCTGAGAGCTTCAACCTGCGGGAGAGCTTCAAGAGCTTCAACCGG 562
 DB 361 CTGCAAGAGGACACCTGCTGAGAGCTTCAACCTGCGGGAGAGCTTCAAGAGCTTCAACCGG 420
 OY 563 CTGCTGGCCATCCCGCGGGCAATCCCGGCTCCCTTCCAGAGAGTGGAGCGCTCATG 622
 DB 421 CTGCTGGCCATCCCGCGGGCAATCCCGGCTCCCTTCCAGAGAGTGGAGCGCTCATG 480
 OY 623 GGAGAGCTGAGTCTGAGAGCTTGGAGAGCCAGCTATGTCGCACTTACAGCTTGGGGGC 682
 DB 481 GGAGAGCTGAGTCTGAGAGCTTGGAGAGCCAGCTATGTCGCACTTACAGCTTGGGGGC 540
 OY 683 ATTTCCAGAGGCTGAGAGCGGGGCTGCTCATCGAGCGCCAGTGTCTCAGAGTCTTAAG 742
 DB 541 ATTTCCAGAGGCTGAGAGCGGGGCTGCTCATCGAGCGCCAGTGTCTCAGAGTCTTAAG 600
 OY 743 GTCAATGCTGTTGATGAGCAACACAGGCTGAGAGCTTCAATGATGCTTCAATGATGTC 802
 DB 601 GTCAATGCTGTTGATGAGCAACACAGGCTGAGAGCTTCAATGATGCTTCAATGATGTC 660
 OY 803 GTCCCTGCTGGTGAAGTGGCTGGAGAAACCAATTTGCTTCAACATTCACAGCCG 862
 DB 661 GTCCCTGCTGGTGAAGTGGCTGGAGAAACCAATTTGCTTCAACATTCACAGCCG 720
 OY 863 GCTTCTGAGCTTTTTCAGCTCTTTTGAACAATTTGAGATCTTCAAGTGGAGAGCTGAT 922
 DB 721 GCTTCTGAGCTTTTTCAGCTCTTTTGAACAATTTGAGATCTTCAAGTGGAGAGCTGAT 780

OY 923 TTCTGTGGACAGCGGAGGAATGCTGATTTCTCAATGATGCTGGTTACCTTGTCT 982
 DB 781 TTCTGTGGACAGCGGAGGAATGCTGATTTCTCAATGATGCTGGTTACCTTGTCT 840
 OY 983 GAACTTCAAAACCTTTTGAAGTCTTATATGAGCTGAGAGTGGATATGCAAGAG 1042
 DB 841 GAACTTCAAAACCTTTTGAAGTCTTATATGAGCTGAGAGTGGATATGCAAGAG 900
 OY 1043 GAACGGGAATAGAAACCTTCCAGAGAGTCCAGATGATGATGATGCTTCCCTACAAATCA 1102
 DB 901 GAACGGGAATAGAAACCTTCCAGAGAGTCCAGATGATGATGATGCTTCCCTACAAATCA 960
 OY 1103 GCAATTTGTATAAATTTTGAAGATATGTAAGATGAACACCTGAAACGTTACCA 1162
 DB 961 GCAATTTGTATAAATTTTGAAGATATGTAAGATGAACACCTGAAACGTTACCA 1020
 OY 1163 ATGGTTCCTTCAAAACCAAGATTTCTCGAGAGTTCCTTAACTGGGCTGCTCTG 1222
 DB 1021 ATGGTTCCTTCAAAACCAAGATTTCTCGAGAGTTCCTTAACTGGGCTGCTCTG 1080
 OY 1223 AGGAGCTGACAAAGAACTTGTGAGAAATTAAGCTGACATGATTAAGCTTCTCTCAG 1282
 DB 1081 AGGAGCTGACAAAGAACTTGTGAGAAATTAAGCTGACATGATTAAGCTTCTCTCAG 1140
 OY 1283 AATCTGATCAATGGGTTTGTCTTCTCTTCTTCTGCTTCTGCGGGTCCGAGCAATGCTCA 1342
 DB 1141 AATCTGATCAATGGGTTTGTCTTCTCTTCTTCTGCTTCTGCGGGTCCGAGCAATGCTCA 1200
 OY 1343 AAGGCTGCTATCCAGAGACCGCTAGGCTCTCTTACAGATTTGGGGGCGCACCCGCTAC 1402
 DB 1201 AAGGCTGCTATCCAGAGACCGCTAGGCTCTCTTACAGATTTGGGGGCGCACCCGCTAC 1260
 OY 1403 ACAGGATGCTGAACGCTGTGAATCTGTTCCGCTGCTGAGAGCTGTCAGCGAGAG 1462
 DB 1261 ACAGGATGCTGAACGCTGTGAATCTGTTCCGCTGCTGAGAGCTGTCAGCGAGAG 1320
 OY 1463 AGTCAGAGAGCGCTTACAGAGAGTGGAGATATGCTGGCTTACATGCACTGACGTCCTC 1522
 DB 1321 AGTCAGAGAGCGCTTACAGAGAGTGGAGATATGCTGGCTTACATGCACTGACGTCCTC 1380
 OY 1523 CCCTTCAAGGTTTGTCCACAGATGATTTTACAGAGTGTGCTGCTGAGAGCTGAGGCTTA 1582
 DB 1381 CCCTTCAAGGTTTGTCCACAGATGATTTTACAGAGTGTGCTGCTGAGAGCTGAGGCTTA 1440
 OY 1583 CATCTGAGAGTGGCCGATTTGGATTTTCTGCTGCTCTTGGCCCCCACTTAAT 1642
 DB 1441 CATCTGAGAGTGGCCGATTTGGATTTTCTGCTGCTCTTGGCCCCCACTTAAT 1500
 OY 1643 GGTGAATTTCTAATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1702
 DB 1501 GGTGAATTTCTAATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
 OY 1703 GTAGTGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1762
 DB 1561 GTAGTGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 OY 1763 CAAGAAATGCCATCTCTTTTAAATCATCAGTATTTTAAATTCACAAATATTCAGT 1822
 DB 1621 CAAGAAATGCCATCTCTTTTAAATCATCAGTATTTTAAATTCACAAATATTCAGT 1680
 OY 1823 GAGATTTCTGTAGTCAATGATGTTTCAAGAGCTGATTTCACTTGGAGCTCAATGTT 1882
 DB 1681 GAGATTTCTGTAGTCAATGATGTTTCAAGAGCTGATTTCACTTGGAGCTCAATGTT 1740
 OY 1883 TCTGTGACAACTAATCAATGCTGCTTCACTCACTCAAGAAATCAATTCATTTAGAAAAC 1942
 DB 1741 TCTGTGACAACTAATCAATGCTGCTTCACTCACTCAAGAAATCAATTCATTTAGAAAAC 1800
 OY 1943 TGCCAGAGTGAACATCTAGATTTCAATGAACCTTCTGATTTTGTATTCATTTTCA 2002
 DB 1801 TGCCAGAGTGAACATCTAGATTTCAATGAACCTTCTGATTTTGTATTCATTTTCA 1860
 OY 2003 GCTCTTGTATCTTGAAGATGTTGTTTCAAAATTAAGGATCATCTCATTAAGCAGTAG 2062

Db	1861	GCCTTGTCTATCTAGCAATAGTTGTTTCAAAATAGGAGTATCTCATTAAGCAGGTAG	1920
RESULT 7			
LOCUS	AY195873		
DEFINITION	Mus musculus strain PERA/EI ATP-binding cassette sub-family G		
ACCESSION	AY195873		
VERSION	AY195873.1	GI:31322257	
KEYWORDS			
SOURCE	Mus musculus (house mouse)		
ORGANISM	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 2351)		
AUTHORS	Wittenburg, H., Lyons, M.A., Li, R., Churchill, G.A., Carey, M.C. and Paigen, B.		
TITLE	Primary Roles of FXR and ABCG5/ABCG8 in Cholesterol Gallstone Susceptibility: Evidence from a Cross of PERA/EI and I/Ln Indbred Mice		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 2351)		
AUTHORS	Lyons, M.A., Wittenburg, H., Walsh, K.A., Carey, M.C. and Paigen, B.		
TITLE	Direct Submission		
JOURNAL	Submitted (11-DEC-2002) The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA		
FEATURES	Location/Qualifiers		
source	1..2351		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="PERA/EI"		
	/db_xref="taxon:10090"		
	/chromosome="17"		
	/map="55 cM"		
	/sex="male"		
	/tissue_type="liver"		
gene	1..2351		
	/gene="Abcg5"		
	139..2097		
CDS	/note="ATP-dependent canalicular cholesterol transporter; white subfamily"		
	/codon_start=1		
	/product="ATP-binding cassette sub-family G member 5"		
	/protein_id="AA045094.1"		
	/db_xref="GI:31322258"		
	/translation="MGEIPELSPEGARPHNIRGSLSLIEGVSVTGEARHSLGLVLAH SYSVNRVGPMMNIRKSCQKMDROILKDVSLYIESGOIMCIISSGSGKTTLLDLSG RLRTGTLEGVEFVNGCELRDQFQDCSYLQSDVFLSLVRETLLRTATLALCRS SADFNKKEAVATELSLHVADOMIGSYNEGISGERRRYSIAOILLDDPKVAMD EPTTGIDCANTAOIVLLAELARDRIVYTHORPSRLFOHFDKIALITLYELVFCG TPBEMLGPPNNKGYPCPEHSNPFDMITSDVDSREHEITVAKVOMLECAFESD ITKILLENIRARYLKTLYPEFKTKDPGGMKGLVLLRRTNLMNRKQAVIRLV OMLNGLEFLIFLHOMNLAVLAVLQVDRGLYQLVGAATPYGMNLNANLEPMLRAVS DOESDGLYHKOMLAVLAVLPEVSATVFSVCWTGLLYEVAERFGFSALV AHLIGELFLIVLIGIVONPNIVNSIVALISGLISGFIKRIQOEPIPIKIIIGYF TPQKCCCELIVNERYNGTSGSNTSLNHPKAIITQGVQFIKTKPGATSRRTAN FLITGSPALVYIIGIVFKVDYILSR"		
BASE COUNT	569 a 605 c 594 g 583 t		
ORIGIN			
Query Match	60.3%; Score 1410.8; DB 10; Length 2351;		
Best Local Similarity	80.4%; Pred. No. 1.4e-296;		
Matches 1665; Conservative	0; Mismatches 402; Indels 3; Gaps 1;		
QY	25	CTGCCAGGGGCTCCCAACTGAGCACTGTGGGAGGGTCCGGCCACACAGAAATTGGC	84
DB	57	CTCCCATTTGGCTCCACAGTTAAAGCTGCCGAGCCGACGACGACAGAAATTTAC	116
QY	85	CCAGCTTGTGCTGCTTGGCCATGAGTACCTTCATCTTTGACCCCGAGGGTCCAT	144
DB	117	TTGCATTTGCTTCTCTGCTAGCCATGGGTGAGTGTCCCTTTCTGAGTCCAGAGGAGCCAG	176
QY	145	GGGTCTCCAAAGTAAACAGAGAGGCTCCAGAGCTCCCGAGGGGGCTCTGCCACGCCCCC	204
DB	177	AGGGCTCTACATCAACAGAGGGGTCTGTAGCTCTCTGGAGCAAGTTCCGTCTCGGGCAC	236
QY	205	GGAGCT---CAAGCTGGGGCATCTCCATGCTCTCTACAGCTGAGCCAGCCGTGAG	261
DB	237	AGAGGCTGGACACAGTTAGTGTGCTGTGATGTCTTACAGGCTGAGCAACCGTGTCCG	296
QY	262	GCCCTGTGGGACATCATCTTTCGGGACAGCAGTGGACAGCAGATCTCAAGATGT	321
DB	297	GCCTGTGGACATCAATCAATGACAGAGAAAGTGGAGAGCAATCTCTCAAGATGT	356
QY	322	CTCCCTGTACGTGGAGGGGGAGATCATATGTGATCTCTAGAGAACTGACGCTCCGGAA	381
DB	357	CTCTGTATCATCGAGAGTGGCCAGTATATGTGATCTTAAAGAGTGTGAGCTCAGGGA	416
QY	382	AACACGCTGTGTGAGCCATGTCTCGGAGGCTGGGGGCGCGGGGACCTCTCTGGGGGA	441
DB	417	GACACGCTGTGTGAGCCATGTCTCGGAGGCTGGGGGCGCGGGGACCTCTCTGGGGGA	476
QY	442	GGTATGTGAAAGCGCGGCGCTGCGCGGGAGCAGTTCCAGAGCTGCTTCTCTACGT	501
DB	477	GGTGTGTGAAAGCGCTGCGCGGGAGCAGTTCCAGAGCTGCTTCTCTACGT	536
QY	502	CGTGCAGAGCAGACCCCTCTGAGCAGCCTCACCGCTGCGGAGAGCTGCTCAGTACCC	561
DB	537	CTGTGAGAGCAGCTTTTCTTGTGAGCAGCTCTGAGCAGCTTGTGAGCAGCTTGTGAG	596
QY	562	GCTGTGAGCAGCTTGTGAGCAGCTTGTGAGCAGCTTGTGAGCAGCTTGTGAGCAGCT	621
DB	597	GATGCTGTGAGCAGCTTGTGAGCAGCTTGTGAGCAGCTTGTGAGCAGCTTGTGAGCAG	656
QY	622	GGCAGAGCTGAGCTGTGAGCAGCTTGTGAGCAGCTTGTGAGCAGCTTGTGAGCAGCT	681
DB	657	GACAGAGCTGAGCTGTGAGCAGCTTGTGAGCAGCTTGTGAGCAGCTTGTGAGCAGCT	716
QY	682	CATTTCCAGAGCTGTGAGCAGCTTGTGAGCAGCTTGTGAGCAGCTTGTGAGCAGCT	741
DB	717	AATTTCCAGAGCTGTGAGCAGCTTGTGAGCAGCTTGTGAGCAGCTTGTGAGCAGCT	776
QY	742	GCTCATGCTGTGTGAGTGAAGCAGCAACACAGAGCTGTGAGTGAAGTGAAGTGAAG	801
DB	777	GGTCAATGATGCTGATGAGCAGCAACACAGAGCTGTGAGTGAAGTGAAGTGAAGTGA	836
QY	802	CGTCTCTGTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA	861
DB	837	CTTCTCTGTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA	896
QY	862	CGTCTCTGTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA	921
DB	897	TGCTCTGTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA	956
QY	922	TTTCTGTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG	981
DB	957	GTTCTGTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG	1016
QY	982	TGACATTCGAAGCTTTGACTTCTATATGAGCTGAGCTGAGTGAAGTGAAGTGAAG	1041
DB	1017	TGACATTCGAAGCTTTGACTTCTATATGAGCTGAGCTGAGTGAAGTGAAGTGAAG	1076
QY	1042	GGAGCGGGAATAGAAAGCTCCAAAGAGTCCAGATGATGATGATGATGATGATGATG	1101
DB	1077	AGAGCGGGAATAGAAAGCTCCAAAGAGTCCAGATGATGATGATGATGATGATGATG	1136
QY	1102	AGCAATTTGCTAATAAAGCTTTGAAAGATGAAAGATGAAAGATGAAAGATGAAAGAT	1161
DB	1137	TGACATTCATCAAAATTTGAGAAACATTTGAAAGATGAAAGATGAAAGATGAAAGAT	1196
QY	1162	AATGTTCTTTTCAAAACCAAGATTTCTGAGATTTTCTTAATCTGAGTGTCTCTCT	1221
DB	1197	CACGTTCTTTTCAAAACCAAGATTTCTGAGATTTTCTTAATCTGAGTGTCTCTCTCT	1256

Db 837 CCTCTCTGCTGAGCTGGCTGCGAGGACCCGAAATGTTCTCATCATCCATCCACCGACC 896
 QY 862 CCGTTCGAGCTTTTTCAGCTCTTTTGACAAATTTGCCATTCCTGAGCTTCGAGAGCTGAT 921
 Db 897 TCCTCTGAGCTTTTTCACACACTTCGACAAATTTGCCATTCCTGAGCTTCGAGAGCTGAT 956
 QY 922 TTTCTGTGCGACGCGACCGGAAATGCTTGAATTTCTTCAATGACTGCGGTTAACCTTGTCC 981
 Db 957 GTTCTGTGCGACCGCGACCGGAAATGCTTGAATTTCTTCAATGACTGCGGTTAACCTTGTCC 1016
 QY 982 TGAACATTTCAACCGCTTTGACTTCTATATGAGCTGACGCTGATGATACCCAAACGAA 1041
 Db 1017 TGAACATTTCAACCGCTTTGACTTCTATATGAGCTGACGCTGATGATACCCAAACGAA 1076
 QY 1042 GGAACGCGAAATGAAACCTTCACAGAGCTCCAGATGATGATGATGATGATGATGATGATGAT 1101
 Db 1077 AGAGCGGAAATGAAACCTTCACAGAGCTCCAGATGATGATGATGATGATGATGATGATGATGAT 1136
 QY 1102 AGCAATTTGCTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1161
 Db 1137 TGAATCTTATCAACAAATTTCTGAGAACTTGAAGAGCAGATGATGATGATGATGATGATGATGAT 1196
 QY 1162 AATGCTTCTTCAACAAATTTCTGAGAACTTGAAGAGCAGATGATGATGATGATGATGATGATGATGAT 1221
 Db 1197 CAGCGCTTCTTCAACAAATTTCTGAGAACTTGAAGAGCAGATGATGATGATGATGATGATGATGATGAT 1256
 QY 1222 GAGAGAGTGCACAGAACTTGTGAGAAATAGCTGAGTATGATGATGATGATGATGATGATGATGATGAT 1281
 Db 1257 GAGCGGAGTGCACAGAACTTGTGAGAAATAGCTGAGTATGATGATGATGATGATGATGATGATGATGAT 1316
 QY 1282 GAATCTGAT 1341
 Db 1317 GAATCTGAT 1376
 QY 1342 AAGGCGCTATCCAGAGCGCTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1401
 Db 1377 AAGGCGCTGATCCAGAGCGCTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1436
 QY 1402 CACAGGAGTGCACAGAACTTGTGAGAAATAGCTGAGTATGATGATGATGATGATGATGATGATGATGAT 1461
 Db 1437 CACGCGAGTGCACAGAACTTGTGAGAAATAGCTGAGTATGATGATGATGATGATGATGATGATGATGAT 1496
 QY 1462 GAGTGCAGAGCGCTTACACAGAGTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1521
 Db 1497 GAGTGCAGAGCGCTTACACAGAGTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1556
 QY 1522 CCCCTTCAGCGCTGATCCAGAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1581
 Db 1557 CCCCTTCAGCGCTGATCCAGAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1616
 QY 1582 AATCTGAGGCTGATCCAGAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1641
 Db 1617 GATCTGAGGCTGATCCAGAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1676
 QY 1642 TGTGTAATTTCTAATCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1701
 Db 1677 TGTGTAATTTCTAATCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1736
 QY 1702 TGTAGGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1761
 Db 1737 TGTAGGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1796
 QY 1762 ACAAGAAATGCCCATCTCTTTAAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1821
 Db 1797 ACAAGAAATGCCCATCTCTTTAAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1856
 QY 1822 TGTAGTCTTGTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1881
 Db 1857 TGTAGTCTTGTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1916
 QY 1882 TTTCTGTGACAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1941
 Db 1917 CTTCTATGCTAATATCACCGCATGCTGCGCCATCACACCGGCGCTGATGATGATGATGATGATGATGATGAT 1976

QY 1942 CTGCCAGAGTGCACATCTAGATTCACAAAGAACTTTCTGATTTGATTCATTTATTC 2001
 Db 1977 CTGCCAGAGTGCACATCTAGATTCACAAAGAACTTTCTGATTTGATTCATTTATTC 2036
 QY 2002 AGCTCTGCTATCTAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2061
 Db 2037 AGCTCTGCTATCTAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2096
 QY 2062 GTCAACCATGCTGCGGAAAGGAAAGTG 2091
 Db 2097 GTTAAAGATGACAGCGAGGAAAGGCTAATG 2126
 RESULT 9
 AF312713 2354 bp mRNA linear ROD 16-MAY-2001
 LOCUS AF312713
 DEFINITION Mus musculus sterolin (Abcg5) mRNA, complete cds.
 ACCESSION AF312713
 VERSION AF312713.2 GI:14091944
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 REFERENCE
 AUTHORS Lee, M.H., Lu, K., Hazard, S., Yu, H., Shulenin, S., Hidaka, H.,
 Kojima, H., Allikmets, R., Sakuma, N., Pegoraro, R., Srivastava, A.K.,
 Salem, G., Dean, M. and Patel, S.B.
 IDENTIFICATION OF A GENE, ABCG5, IMPORTANT IN THE REGULATION OF
 DIETARY CHOLESTEROL ABSORPTION
 Nat. Genet. 27 (1), 79-83 (2001)
 JOURNAL 20578753
 MEDLINE 11138003
 PUBMED 2 (bases 1 to 2354)
 REFERENCE Lu, K., Lee, M.-H. and Patel, S.B.
 AUTHORS Direct Submission
 JOURNAL Submitted (12-OCT-2000) Division of Endocrinology, Diabetes and
 Medical Genetics, Medical University of South Carolina, 114 Doughty
 St, STB 541, Charleston, SC 29403, USA
 TITLE 3 (bases 1 to 2354)
 REFERENCE Lu, K., Lee, M.-H. and Patel, S.B.
 AUTHORS Direct Submission
 JOURNAL Submitted (16-MAY-2001) Division of Endocrinology, Diabetes and
 Medical Genetics, Medical University of South Carolina, 114 Doughty
 St, STB 541, Charleston, SC 29403, USA
 REMARK
 COMMENT Sequence update by submitter
 FEATURES
 ON May 16, 2001 this sequence version replaced gi:12382299.
 location/Qualifiers
 1..2354
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /tissue_type="liver"
 1..2354
 /gene="Abcg5"
 139..2097
 /gene="Abcg5"
 /note="ABCG5"
 /codon_start=1
 /product="sterolin"
 /protein_id="AA653097.1"
 /db_xref="GI:12382300"
 /translation="MEGLPELSPEGARPHINRSGLSLSDGQSVGTENRHSGLVLAH
 SYSVSRKVPWNINISQDQKMDKDVSLYIESQIMCIISSGSGKRTLLDAISG
 RLRRCTLEGEEVFNCGELRPODFCVISVDSVLSLVRRTLRMTALGCRS
 SADENKKEAVMTSELSSHVADOMGNSNFGSGISGERRRVSIAQLQDPRVAMLD
 EPTGDKMTANQIYLLAELARRBRIYVLTTHORSELPQFQDKAIITITGELVFCG
 TPBEMLGFNNKGYRPOEHSNPFDFMDLTSVDTGSREREIEFYRVMQLKCAFESE
 IYKILLENIERARYLKTPTLPKTRKDPGKGLGVGLRYRTNRMKQKQVINYHLV
 ONLNGELFIYLIRYONNTLKGAVODRGLIYOLGVGATPYGMNANVLAIFNLRAVS
 DOESDGIYHKQMMLAIVLAVLIPRSVIATVIFSSVCWTGLLVEVNAFQFSNALL

Mon Jul 28 09:43:27 2003

us-09-989-981a-5.rge

Page 13

BASE COUNT 573 a 604 c 594 g 583 t
ORIGIN
A P H I G E E F I L V L I G I V O A N N I N V A S I L L I S G S G F I R I O E M P I P K I L G F
T P R K C C E I L V N E F V E G L N F T G G S N S I M L N H P R C A I T O G V Q I E K T C P G A T S E T A N
F L I L G S I P A L V I L G I Y I F R V D I L I S R "

Query Match	60.24;	Score 1409.2;	DB 10;	Length 2354;
Best Local Similarity	80.48;	Pred. No. 3e+296;		
Matches 1664;	Conservative	0;	Mismatches 403;	Indels 3;
				Gaps 1;

[illegible]

QY	982	TGAACATTTAAACCCCTTTTGA	CTCTATATAGACCTGACGTC	CACTGATGATACCCAAACAA	1041
Db	1017	TGACATTTCCAAATCCCTTGTG	ATTTTATCATGACTGCAATCA	CTGACACCTGGCACCCAAACAG	1076
QY	1042	GGAAACGGAAATAGAANAACCT	CCACAGAGAGTCACAGATGAT	TAGATATCTGCTTCAGAAATC	1101
Db	1077	AGAGCGGGAAATATAGAAAGST	ACACACGAGTACAGATGCTGA	ATATGTGCTTTCAGAGATC	1136
QY	1102	AGCAATTTGCTATAAACTTGTG	AAGATATTTGAAAAGATGAAAC	ACCTGAAAACGTTACC	1161
Db	1137	TCACATCTATCAAAAATTTCTG	SAGAACATTTGAAAAGACACG	ATACCTGAAAACCTTACC	1196
QY	1162	AAATGGTCCCTTCAAAACCAAA	AGATTTCCCTGAGCTTTTCTTA	ACTGGGTCTTCTCCT	1221
Db	1197	CACGGTCTCTTCAAAACCAAA	AGATTCCTCGGAGATGTTTGG	CAAAAGCTTGGTCTCCTCT	1256
QY	1222	GAGAGAGTGAACAGAAACTTGTG	TAGAAAATACCTGGACAGTAT	TACAGCGTCCCTCA	1281
Db	1257	GAGCGAGATTAACAGAAACTTA	TATAGAGAAATACAGAGCA	TGATATATGTGCTTCCTCA	1316
QY	1282	GAAATCTGATCATGGCTTTGTT	CTTCCTCTTCTCGTCTCGGAG	TCCGAAACAATGTGCT	1341
Db	1317	GAAATCTGATCATGGGCTCTTC	CTCCTCATTTTCTACCTTCT	CCGCGCTCCAGAAACAACAG	1376
QY	1342	AAAGGCTGTATCCAGAGACGGG	AGAGTCTCCTTTACCAATTTT	GTGGGCGCCACCCCTGA	1401
Db	1377	AAAGGCGCTGTGCAAGAGACCG	GTGGGCTCCTATACAGTTGTG	GTGGTGCACCCCTATA	1436
QY	1402	CACAGCAGTGTGAACCGTGTGA	ATCTGTTCCCGTCTGAGACTG	CTCAGCCACCAAGA	1461
Db	1437	CACGGCATGTGTCATGTGTGA	ATCTGTATCCATCTGAGACCG	CTCAGCCACCAAGA	1496
QY	1462	GAGTACAGACGCGCTCTACCA	GAAGTGGCAGAGATGCTGAC	CTATAGCACTGCACGTCCT	1521
Db	1497	GAGTCAGAGATGGCCGTGTAT	CAATTAATGTGGACAGATGCT	CTCGCCCTACGTGCTACAC	1556
QY	1522	CCCCCTTACGCGTTGTGGCACA	CAATTTTTCACAGAGTGTGCT	AGACTGAGACGCTGGGCTT	1581
Db	1557	CCCCCTTACGCGTATCGCACAG	TCATGGTCATTTTTCAGCAGT	GTGTATTTTGGACTCTGGGCTT	1616
QY	1582	ACATCCTAGAGTCCCGGATTTG	AGATATTTTTCGCTGCTCTCT	GTGGCCCCCCCACCTAAT	1641
Db	1617	GTATCCGTAAATGTGCAGATTT	GGATATTTTCTGTGCTCTTTT	GGCCCCCTCACCTAAT	1676
QY	1642	TGTGGAATTTTAACTCTTGTGT	ACTTGTGTATGTCGCAAAATC	CAAAATATATGCTCAACAG	1701
Db	1677	TGGAGAAATTTTAACTTGTGTG	CTGTGCTTGGATATGTCGCA	AAACCCATAATATGTCAACAG	1736
QY	1702	TGTATGGCTCTGTGTGCCAAT	TGGGGGGGCTGTGTGATCTG	GTGCAATTCCTCAGAAACAT	1761
Db	1737	TATATGGCTCTGTGTGCCAAT	CTCTGTGGCGCTTATTTGGA	TATCGGATTTATATCAAAACAT	1796
QY	1762	ACAAGAAATGCCCATTCCTTTT	AAAAATCATCAGTATATTTTA	CATTCGCAAAAATATATGCAAG	1821
Db	1797	ACAAGAAATGCCCATTCCTTTT	AAAAATCATCAGTATATTTTA	CATTCGCAAAAATATCTGTG	1856
QY	1822	TGAATTTCTTGTACTCAATGAT	TTCTACGAGATTTTCACTTGTG	CGAGCTCAAAATGT	1881
Db	1857	TGAATTTCTTGTGTACTCAATG	ATTTTACGAGATTTTCACTTGTG	CGAGCTCAAAATGT	1916
QY	1882	TTTGTGTACACTTAATCCAAAT	TGTCCTTCACTCAAGAAATTC	CACTCAAGAAATTC	1941
Db	1917	CTTATCTTAATACCCGATATG	CGCCATACCCCAAGGGGTCC	CACTTATGAGAAAC	1976
QY	1942	CTGCCAGGTGCAACATCTAGAT	TTCACAAATGAACTTCTGAT	TTTGTATATCAATTTATCC	2001
Db	1977	CTGCCAGGTGCTACATCCAGAT	TTCACCGGAAATCTTCATCT	TATATGGTTTATCC	2036
QY	2002	AGCTCTTGTATCCTAGGAAT	ATAGTGTTTTCAAAATATAGG	ATATCACTTACAGAGTA	2061
Db	2037	AGCTCTTGTATCCTATGGAAT	ATAGTGTTTTAAAGTCAGGAG	CTACCTATATAGCAGATA	2096

QY 2062 GTGAAGCATGCTGGGAAAATGAGATG 2091
 DB 2097 GTTAAGATGACGACGAGGAGGTTATG 2126

RESULT 10 AY195872

LOCUS 2351 bp mRNA linear ROD 01-JUN-2003
 DEFINITION Mus musculus strain I/LmJ ATP-binding cassette sub-family G member
 5 (Abcg5) mRNA, complete cds.
 ACCESSION AY195872
 VERSION AY195872.1 GI:31322255

KEYWORDS

SOURCE

Mus musculus (house mouse)

REFERENCE

1 (bases 1 to 2351)
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Wittenburg, H., Lyons, M.A., Li, R., Churchill, G.A., Carey, M.C. and
 Paigen, B.

TITLE

Primary Roles of FXR and ABCG5/ABCG8 in Cholesterol Gallstone
 Susceptibility: Evidence from a Cross of PERA/EI and I/Lm Inbred
 Mice

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 2351)
 Lyons, M.A., Wittenburg, H., Walsh, K.A., Carey, M.C. and Paigen, B.
 Direct Submission
 Submitted (11-DEC-2002) The Jackson Laboratory, 600 Main Street,
 Bar Harbor, ME 04609, USA

JOURNAL

Location/Qualifiers

FEATURES

source

1.2351
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="I/LmJ"
 /db_xref="taxon:10090"
 /map="55 CM"
 /sex="male"
 /tissue_type="liver"

1.2351

/gene="Abcg5"

139..2097

/gene="Abcg5"

/note="ATP-dependent canalicular cholesterol transporter;
 white subfamily"

/codon_start=1

/product="ATP-binding cassette sub-family G member 5"

/protein_id="AA045093.1"

/db_xref="GI:31322255"

/translation="MGELPFLSPEGARPHINRGSLSLEQSVTGTARHSLVILH
 SYSNRYGPMWNKSCQKMDROLKRVSLITSGQIMCIISSGSGKTYLLDLSIG
 SADFTNKVEAVMTSLSHVADYIGSVLQSDVFLSLVRETLRTALALCRS
 EPTTGIDMTANQIVLLAEIARDRIVYTHOPRSELFQHFDAITLITGELVFCG
 TPEDMGFFNNQYPCPEHNPFDYMDLOSREIRETYRYVOMLESAFESD
 IYKILLENIERARYKTLPTVPEFKRDPGMRGKLVILRRTRNLMNRKQAVIRLV
 DONLIMGFLFIYLLRVONNTLKAGVQDVGILYQVGAATPYGMNANVLEPMLKAVS
 DOESODGLYHNMKMLATVLAHPISIIATVIFSSVCWTGLYEVARFGFSAAL
 APLHIGELTIVLVLGIVONPNIVSIVALLSGLISGIRINIOEMPIKILGYE
 TFOKYCCELIVNEFYELFTGGSNTMLNHPKCAITQGVETFEKTEPGATSRRTAN
 FLIYGFPAIVIIIGIVIFKVRDYILSR"

BASE COUNT 577 a 598 c 590 g 586 t

ORIGIN

Query Match

Best Local Similarity 60.1%; Score 1406; DB 10; Length 2351;
 Matches 1662; Conservative 0; Mismatches 405; Indels 3; Gaps 1;

QY 25 CTGCACAGGCTCCCAACAGCACTCTGGGAGGCTCGGACACAGAAATTTGC 84
 DB 57 CTCACATGGCTCTCTCACTTAAGCTGCGCGAGCGGACAGGCACTAGAAATATCAC 116
 QY 85 CCAAGCTTGCTGCTGTGGCATGTGAGCTCTCATCTTTGACCCCGAGGCTCAT 144

DB 117 TTGCATTTGCTCCGCTAGCCATGGTGGAGTCCCTTCTAGATCCAGAGGACCCAG 176
 QY 145 GGGTCTCCAACTTAACAGAGCTCCAGAGCTCCCTGGAGGGGGCTCTCCACGCCCC 204
 DB 177 AGGGCTCTACATCAACAGAGGGTCTCTGAGCTCCCTGGAGGAGGTTGGTACAGGGGAC 236
 QY 205 GGAGCCT---CAGAGCCCGGCAATCCATCCATCCCTCTACAGCGCTACAGCCCGGTAG 261
 DB 237 GGAGGCTCGGACAGCTTGGTGTCTGCTGATGTCTCTTCCAGCGCTCAGCAACCTGTGCG 296
 QY 262 GCCCTGTGGACATCACTATCTTCCGCGACAGTGGAGCACGGGAGATCCCAAGAGT 321
 DB 297 GCTTGGTGGACATCAATATCATTCAGAGCAAGATGGAGACGGCAATCTCAAGAGT 356
 QY 322 CTCTTTAGCTGAGAGCGGGGAGATCATGTGATCTTACAGAGCTCAGGCTCGGGAA 381
 DB 357 CTCTTTAGCTGAGAGCGGGGAGATCATGTGATCTTACAGAGCTCAGGCTCAGGAA 416
 QY 382 AACACGCTCTGGAGCCCAATGTCGGGAGGCTGGGGCGGGGACCTTCTCGGGGA 441
 DB 417 GACACGCTCTGGAGCCCAATGTCGGGAGGCTGGGGCGGGGACCTTCTCGGGGA 476
 QY 442 GGTGTATGTAACGGCGCGGCGCTGCGCGGAGACATTCAGAGACTCTTCTCTAGT 501
 DB 477 CGTGTATTGAATGGCTGGAGCTGCCAGAGGACCACTTCAAGACTCTTCTCTAGT 536
 QY 502 CTGCAAGAGCAACCTGCTGAGACAGCTTACCGCGGAGAGCGTCAACACCG 561
 DB 537 CTGCAAGAGCAACCTTCTGAGACAGCTTACCGCGGAGAGCGTCAACACCG 596
 QY 562 GCTGCGCCATCGCGCGGCGCAATCCCGCTCTCTCAAGAGAGTGAGGCGCTCAT 621
 DB 597 GATGCTGGCCCTTGGCGGAGCTCGCGGACTTCTCAACAGAGTGAAGGACGATG 656
 QY 622 GGCAGAGCTGAGTCTGAGCATGTGACAGACCGAGTATGGCACTACAGCTGGGGG 681
 DB 657 GACAGAGCTGAGCTGAGCATGTGACAGACCGAGTATGGCACTACAGCTGGGGG 716
 QY 682 CATTCACAGGCTGAGCGCGCGGCTGTCATCGGACCCAGCTCTCAGGATCTTA 741
 DB 717 AATTCAGTGGCGAGCGCGCGGCTGTCATCGGACCCAGCTCTCAGGATCTTA 776
 QY 742 GGTATGCTGTTGATGAGCAACCAAGAGCTGAGCTGATGATGATGATGAT 801
 DB 777 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 836
 QY 802 GCTCTCTGCTGAGCACTGCTGAGCAAGCAATGCTGCTTCCACCATCACCAGC 861
 DB 837 CTTCTCTGCTGAGCACTGCTGAGCAAGCAATGCTGCTTCCACCATCACCAGC 896
 QY 862 CCGTCTGAGCTTTTACAGCTCTTGCACAAATTCATCTGAGCTTCCGAGAGTGT 921
 DB 897 TCGCTGAGCTTTTACAGCTCTTGCACAAATTCATCTGAGCTTCCGAGAGTGT 956
 QY 922 TTTCTGAGCACTGAGCAAGCAATGCTGATTTCTCAATGATGCTGCTGCTGCT 981
 DB 957 GTTCTGAGCACTGAGCAAGCAATGCTGATTTCTCAATGATGCTGCTGCTGCT 1016
 QY 982 TGAACATTCAAACCTTTGACTTATATGAGCTGAGCTGAGCTGAGCTGAGCT 1041
 DB 1017 TGAACATTCAAACCTTTGACTTATATGAGCTGAGCTGAGCTGAGCTGAGCT 1076
 QY 1042 GGAACGGAATAGAACTTCAAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1101
 DB 1077 AGACCGGGAATATCAACATCAAGCGAGTACAGTGTGGAATCTGCTTCAAGATC 1136
 QY 1102 AGCAATTTGATTAACCTTTGAAGATATGGAAGATGAACACCTGAAAGCTTACC 1161
 DB 1137 TGACATCTATCAAAATTTGGAAGATATGGAAGATGAACACCTGAAAGCTTACC 1196
 QY 1162 AATGCTCTTCAAAACCAAGATTTCTGAGTGTCTTCAATGAGGCTGCTTCT 1221

D	1197	CACGGTTCCTTTAAACAAAGATCCTCTGGAGTGTTCGGCAAGCTTGTGCTCT	1256
O	1222	GAGGAGATGACAGAAAGATCTGTGAGAAATAGCTGACATGATTTACGCTCTCTCA	1281
D	1257	GAGCGAGTACAGAAAGATCTGTGAGAAATAGCTGACATGATTTACGCTCTCTCA	1316
O	1282	GAACTGATCATGGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1341
D	1317	GAACTGATCATGGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1376
O	1342	AAAGGTCCTATTCAGAGACCGGCTAGTCTCTCTCTCTCTCTCTCTCTCTCTCT	1401
D	1377	AAAGGTCCTATTCAGAGACCGGCTAGTCTCTCTCTCTCTCTCTCTCTCTCTCT	1436
O	1402	CACAGGATGCTGAGACCGGCTAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1461
D	1437	CACAGGATGCTGAGACCGGCTAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1496
O	1462	GAGTCAGAGAGGCTCTTACAGAAAGTGGCAGATGATGCTGAGTACGACGCTCT	1521
D	1497	GAGTCAGAGAGGCTCTTACAGAAAGTGGCAGATGATGCTGAGTACGACGCTCT	1556
O	1522	CCCCCTCAGGCTGTGCGACACAGATTTTTCAGAGTGTGCTGCTGCTGCTGCT	1581
D	1557	CCCCCTCAGGCTGTGCGACACAGATTTTTCAGAGTGTGCTGCTGCTGCTGCT	1616
O	1582	ACATCTGAGGCTGTGCGACACAGATTTTTCAGAGTGTGCTGCTGCTGCTGCT	1641
D	1617	ACATCTGAGGCTGTGCGACACAGATTTTTCAGAGTGTGCTGCTGCTGCTGCT	1676
O	1642	TGCTGATTTTCTAATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1701
D	1677	TGCTGATTTTCTAATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1736
O	1702	TGCTGATTTTCTAATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1761
D	1737	TGCTGATTTTCTAATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1796
O	1762	ACAAGAAATCCCATCTCTTTTAAATCATCATGATTTTAAATCATCATGATTTT	1821
D	1797	ACAAGAAATCCCATCTCTTTTAAATCATCATGATTTTAAATCATCATGATTTT	1856
O	1822	TGAGATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1881
D	1857	TGAGATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1916
O	1882	TGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1941
D	1917	TGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1976
O	1942	CTGCGCAGGTCAGAAATCTGATGATGATGATGATGATGATGATGATGATGATG	2001
D	1977	CTGCGCAGGTCAGAAATCTGATGATGATGATGATGATGATGATGATGATGATG	2036
O	2002	AGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2061
D	2037	AGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2096
O	2062	GTCGAAAGCATGCTGGGAAATGGAATG 2091	
D	2097	GTTAAGATGACAGCAGGAAAGGTTAATG 2126	

RESULT 11
AX320881 2258 bp DNA linear PAT 14-DEC-2001
DEFINITION Sequence 2 from Patent W0019272.
ACCESSION AX320881
VERSION AX320881.1 GI:17902431
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	1	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	Tian, H., Schultze, J. and Shan, B.	
TITLE	Sitosterolemia susceptibility gene (ssg): compositions and methods of use	
JOURNAL	Patent: WO 019272-A 2 25-0CT-2001;	
FEATURES	Tularik Inc. (US)	
source	Location/Qualifiers	
	1..2258	
	/organism="Mus musculus"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:10090"	
	/note="mouse sitosterolemia susceptibility gene (SSG)"	
	47..2005	
	/note="mouse sitosterolemia susceptibility gene (SSG)"	
	/protein	
	/codon_start=1	
	/protein_id="CAD19408.1"	
	/db_xref="GI:17902437"	
	/translation="MGEPLPESPEGAPRHNRGSLSLSDGVTGTEARSLCYLHV	
	SYVSNRVGPMWNKSCQKMDROILKDVSLYISGIMCTLSSGSGSKTLDAISG	
	RLRTGTEVEFVNGEELRDQDQDESVYLDVDELSLVEETLRYTLALCLRS	
	SADYFNKRVAVMTELSLHVADQMSYNGISGERRRYSIAAOLLOPKYMDL	
	EPTGLDQMTANOIVLLAEIARDRIVYIHPHSEELPHQPKIALITGELVFCG	
	TPREMLGFNNCGYRPEHNSRPPVDLTSVDQSREREIETRYKRVOMECAKRES	
	IYHILNIEERARKTLDPVPEFTKPPGPFGLVLRVTLNLRNKAIVMLV	
	QNLIMGLLFLYLRVONNLKGVADRVGLLQDVATPYTGMNLNVAIPMLRAVS	
	DOESQDLFHRKQMLLAVYHLVPEFSAVYAFVSSVYGLGLPEVAREFYSALL	
	APHILGELFLVVLGIYONPNIVISYVALLISGLSGFIRNIOEMPFLKILGF	
	TPQYCCELIVNEEYGLNFTGGSNYSMLNHPICALITQGVQFIERKCPATSRFTAN	
	FLIYGTPLPLVIGIYIFEVRYLYSR"	
BASE COUNT	549 a 579 c 567 g 563 t	
ORIGIN		
Query Match	59.6%; Score 1395.6; DB 6; Length 2258;	
Best Local Similarity	80.7%; Pred. No. 2.8e-293;	
Matches 1642; Conservative	0; Mismatches 389; Indels 3; Gaps 1;	
O	61	GGGTCCGGCCACAGAAATTTGGCCAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCT
D	1	GGGACAGGCCACAGAAATTTCAATTTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT
O	121	ATCTTGACCCCGGAGGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
D	61	CTTCTGATGTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
O	181	GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
D	121	GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
O	238	CTACAGGTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
D	181	CTACAGGTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
O	298	GACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
D	241	GACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
O	358	CCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
D	301	CTTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
O	418	GCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
D	361	GCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
O	478	GTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
D	421	GTTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
O	538	GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
D	481	GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG

OY 598 CCAGAGAGAGTGGAGGCCGTCATGCGAGAGCTAGTCTGAGCCATGTGCGAGACCGACT 657
 DB 541 CAAAGAGAGAGTGGAGGCCGTCATGCGAGAGCTAGTCTGAGCCATGTGCGAGACCGACT 600
 OY 658 GATTGGACACTACAGCTTTGGGGGGCATTTCCAGGGGAGAGGGGGCCGGGCTTCATCCG 717
 DB 601 GATTGGACACTATATATTTGGGGGAAATTTCCAGTGGGAGGGGGCCGGGCTTCATCCG 660
 OY 718 AGCCAGACTCTCCAGAGATCTTAAGGTGATGCTGTTTGAATGAGCCAAACAGAGCCGGA 777
 DB 661 AGCCAGACTCTCCAGAGATCTTAAGGTGATGCTGTTTGAATGAGCCAAACAGAGACTGGA 720
 OY 778 CTGATGATCTGCTAATGATTTGTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 837
 DB 721 CTGATGATCTGCTAATGATTTGTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
 OY 838 TGTGATCTGCTAATGATTTGTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 897
 DB 781 TGTGATCTGCTAATGATTTGTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
 OY 898 CATCTGAGCTGCTGAGAGCTGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 957
 DB 841 CATCTGAGCTGCTGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
 OY 958 CAATGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1017
 DB 901 CAATGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
 OY 1018 GAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1077
 DB 961 GAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 OY 1078 GATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1137
 DB 1021 GATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
 OY 1138 AATGAAACACTGTAAGAAAGCTTACCAATGCTTCTTCAAAACCAAGATTCCTGAGAT 1197
 DB 1081 AGCAACACTGTAAGAAAGCTTACCAATGCTTCTTCAAAACCAAGATTCCTGAGAT 1140
 OY 1198 TTTCTTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1257
 DB 1141 GTTGGGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
 OY 1258 GGCAGTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1317
 DB 1201 GGCAGTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
 OY 1318 TCTGCGGCTCCAGAGATGCTGCTAAGGGTCTATCCAGAGACCGGTAGTCTCCTTA 1377
 DB 1261 TCTGCGGCTCCAGAGATGCTGCTAAGGGTCTATCCAGAGACCGGTAGTCTCCTTA 1320
 OY 1378 CCAAGTTTGGGGGGCCACCCGTCAGAGATGCTGAAAGCTGTGAATCTGTTCCCGT 1437
 DB 1321 TCAAGTTTGGGGGGCCACCCGTCAGAGATGCTGAAAGCTGTGAATCTGTTCCCGT 1380
 OY 1438 GCTGCGAGCTGCTGAGGAGCAGAGAGATGAGAGCGCTTCCAGAACTGAGAGATGAT 1497
 DB 1381 GCTGCGAGCTGCTGAGGAGCAGAGAGATGAGAGCGCTTCCAGAACTGAGAGATGAT 1440
 OY 1498 GCTGCGAGCTGCTGAGGAGCAGAGATGAGAGCGCTTCCAGAACTGAGAGATGAT 1557
 DB 1441 GCTGCGAGCTGCTGAGGAGCAGAGATGAGAGCGCTTCCAGAACTGAGAGATGAT 1500
 OY 1558 TGTGCTGCTGAGGAGCAGAGATGAGAGCGCTTCCAGAACTGAGAGATGAT 1617
 DB 1501 TGTGCTGCTGAGGAGCAGAGATGAGAGCGCTTCCAGAACTGAGAGATGAT 1560
 OY 1618 TGTGCTGCTGAGGAGCAGAGATGAGAGCGCTTCCAGAACTGAGAGATGAT 1677
 DB 1561 TGTGCTGCTGAGGAGCAGAGATGAGAGCGCTTCCAGAACTGAGAGATGAT 1620

OY 1678 CCAAAATCAAAATATAGTCAACAGTATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1737
 DB 1621 CCAAAATCAAAATATATTTTCAACAGTATATAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
 OY 1738 TGTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1797
 DB 1681 TGTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
 OY 1798 TTTTACATTTCAAAATATATGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1857
 DB 1741 TTTTACATTTCAAAATATATGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
 OY 1858 TTTTACATTTCAAAATATATGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1917
 DB 1801 TTTTACATTTCAAAATATATGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
 OY 1918 AGGATCAATCAATGAG 1977
 DB 1861 AGGATCAATCAATGAG 1920
 OY 1978 TGTGATTTTGTATGATTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2037
 DB 1921 TGTGATTTTGTATGATTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
 OY 2038 AAGGATCAATCAATGAG 2091
 DB 1981 AAGGATCAATCAATGAG 2034

RESULT 12
 AF312714
 LOCUS
 DEFINITION
 Rattus norvegicus sterol (Abcg5) mRNA, complete cds.
 AF312714
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Rattus norvegicus (Norway rat)
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 2470)
 Lee, M.H., Lu, K., Hazard, S., Yu, H., Shulenin, S., Hilde, H.,
 Kojima, H., Allikmets, R., Sakuma, N., Pegoraro, R., Silvestra, A.K.,
 Salen, G., Dean, M. and Patel, S.B.
 Identification of a gene, ABCG5, important in the regulation of
 dietary cholesterol absorption
 Nat. Genet. 27 (1), 79-83 (2001)
 2 (bases 1 to 2470)
 Lu, K., Lee, M.-H. and Patel, S.B.
 Direct Submission
 Submitted (12-OCT-2000) Division of Endocrinology, Diabetes and
 Medical Genetics, Medical University of South Carolina, 114 Doughty
 St, STB 541, Charleston, SC 29403, USA
 3 (bases 1 to 2470)
 Lu, K., Lee, M.-H. and Patel, S.B.
 Direct Submission
 Submitted (16-MAY-2001) Division of Endocrinology, Diabetes and
 Medical Genetics, Medical University of South Carolina, 114 Doughty
 St, STB 541, Charleston, SC 29403, USA
 4 (bases 1 to 2470)
 Sequence update by submitter
 Lu, K., Lee, M. and Patel, S.B.
 Direct Submission
 Submitted (26-AUG-2002) Division of Endocrinology, Diabetes and
 Medical Genetics, Medical University of South Carolina, 114 Doughty
 St, STB 541, Charleston, SC 29403, USA
 REMARK
 COMMENT
 On Aug 26, 2002 this sequence version replaced gi:14091945.
 FEATURES
 location/Qualifiers
 1..2470
 source

organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="liver"
1..2470
/gene="Abcg5"
65..2023
/gene="Abcg5"
/note="ABCG5"
/codon_start=1
/product="sterolin"
/protein_id="AAG35098.3"
/db_xref="GI:22477144"
/translation="MSLEPLSPGARGPHNNRGSOSLSLEGVTSYSEARHSLGVANY
SFVSNNRGLVPMNNIKSCQKDRKILKDVSLYIESGOTMCTIGSSGSKTLLDAISC
RLRRGTLEGEVFNCGELRRDOPDCVSYLSDPLSLTVRETLRYTLMALRRSS
SADRYDKVEAVLELSISYADOMIGNYNGGISSEGRVSTDAOLLDPKVMILDS
EPTGLDQMTNHHVLLVELARNRVIVTHIOPRSELEHHPKIALITGEVFCG
TPRMLGPFNNCGTPOPEHSPNDEYMDLSVDYOSREIRELEYKRVOMLESARFSD
ICHKILNLEIETRLKTLPMVPEKTPMPCFKLGVLRVETNLMNRKOVYIMRLV
ONLNGLEIETRLKTLPMVPEKTPMPCFKLGVLRVETNLMNRKOVYIMRLV
DOESODIYOKOMOLAVLHALPESTIVATYIESVCGYKTLGLYEPVAREGESAL
APHLIGELIVLIGVONPRIVNSIVALLISGLIGSGEIRNIEEMPILKILIGYE
TFQKYCCETLIVNEFYGLNFTCGSNTSVPNPMSMTGIOFIETKPCGATSRFTTN
FLIISTIPFLVILGVNFEVRYDLISR"

BASE COUNT 609 a 631 c 623 g 607 t
ORIGIN

Query Match 59.1% Score 1383.8; DB 10; Length 2470;
Best Local Similarity 80.0% Pred. No. 1e-290;
Matches 1641; Conservative 0; Mismatches 407; Indels 3; Gaps 1;

44 TGAAGCCACTCTGGGAGGCTCCGCGCCAGCAAGAAATTTGCCCGCTTGTGCGCTGTG 103
104 GCCATGGGTGACCTCTCATCTTTTGACCCCGGAGGGTCATGGGCTTCCAAGTAAACAGA 163
62 GCCATGATGAGCTGCCCTTCTGAGTCAGAGGAGGACCAAGAGGGCTCACAACACACAGA 121
164 GGGCCAGAGACTCCCGGAGGAGGGGCTCCGCCACCGCCCGGAGCT---CACAGCGTG 220
122 GGGTCTCAGAGCTCCCGGAGGAGGCTCAGTACAGGCTCAGAGGCTCGGACAGCTTA 181
221 GGCATCCTCCATGCTCTCTACAGCGTCAGCGGCTGAGGCGCTGGTGGGACATACAGA 280
182 GAGTCTCTGAAATGTGTCTTACAGCGTCAGAACCGTGTGGGCGCTGGTGAACATCAAA 241
281 TCTTTCGCGGAGCAGTGGAGCAGGACAGATCCTCAAAAGTCTCTCTTGTACGTGGAGAG 340
242 TCAATGCCAGCAGAGTGGAGCAGGAAATCTCAAAAGTCTCTCTTGTACGTGGAGAGT 301
341 GGGCAGATGATGTGCAATCTCTAGAGACTCAGGCTCCGGGAAAACACGCTGTGGAGCGC 400
302 GGCAGACCAATGTGCAATCTCTAGAGACTCAGGCTCCGGGAAAACACGCTGTGGAGCGC 361
401 ATGTCCGGAGAGCTGGGGGCGGCGGAGGACCTTCTGGGGAGGTGTATGTGAACGGCGCG 460
362 ATCTCTGGAGAGCTGGGGGCGGCGGAGGACCTTGGAAAGGTGTATGTGAACGGCGCG 421
461 GCGCTCGCGGAGAGCTTCCAGAGACTCTTCTCTTACGCTCGAGAGCAGCAGCTG 520
422 GAGCTCGGAGGAGCAGTTCAGAGCTGCTCTCTTCTCTCTCTCTCTCTCTCTCTCTCT 481
521 CTGAGAGGCTCAGCCTGCGGAGAGCGCTGACACTACCGCGCTGTGGCAATCCGCGCG 580
482 CTGAGAGGCTCAGCCTGCGGAGAGCGCTGAGATACACGCGATGCTGTGCTCTCCGACAG 541
581 GGCAATCCCGGCTCTCTCCAGAGAGAGTGGAGGCGCTGATAGCAGAGCTGAGTCTGAGC 640
542 AGCTCCGGGAGCTTCTAGACAGAGAGTGGAGGAGCTCTGACAGAGCTGAGTCTGAGC 601

641 CATGGGAGAGCCGAGCTGATGGCAACTACAGCTTGGGGGCGATTTCCAGGGGTAGCGG 700
602 CAGGTGGAGACCAATATATATGCGCAACTATATTTTGGGGGATTTCCAGTGGCAGCGG 661
701 GCGCGGCTCTCCATGCGCAGCCAGCTGCTCCGAGATCTGAAGGATCTGCTTATATAG 760
662 CGCGAGGTCTCATGCGCAGCCCACTCTTACAGGCCCAAGGTCTATATCTGTAGAG 721
761 CCAACACAGGCTGAGCTGATGATGCTATATGATGATGCTGCTGCTGCTGCTGCTG 820
722 CCAACACAGGCTGAGCTGATGATGCTATATGATGATGCTGCTGCTGCTGCTGCTG 781
821 GCTGCGAGGAGCCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 880
782 GCTGCGAGGAGCCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 841
881 CTCTTTGACAAATTTGCCATCTGAGCTGCGAGAGAGCTGATTTCTGCGACGCGACG 940
842 CACTTCGACAAATTTGCCATCTGAGCTGCGAGAGAGCTGATTTCTGCGACGCGACG 901
941 GAAATGCTTGAATTTCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1000
902 GAGATGCTGCGCTTCTTCAATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 961
1001 GACTTCTATATGAGACCTGAGCTGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1060
962 GATTCTATATGAGACCTGAGCTGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1021
1061 TCCAAGAGAGCTCAGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1120
1022 TACAAGAGAGCTCAGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1081
1121 TTGAAGATATTTGAAGATGAAACACCTGTAACCAATGCTTCTTCAAAAC 1180
1082 CTGAGAGCAATTTGAAGATGAAACACCTGTAACCAATGCTTCTTCAAAAC 1141
1181 AAGATCTCTCTGAGCTTCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1240
1142 AAAATCTCTCTGAGCTTCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1201
1241 TTGGTGAATATGATGCTGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1300
1202 CTATATGAGATATGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1261
1301 TTCT 1360
1262 TTCT 1321
1361 GCGGTAGGCTCTCTTACCAAGTTTGTGGGCGCCACCGCTGACACAGGATGTGAACGCT 1420
1322 GCGGTAGGCTCTCTTACCAAGTTTGTGGGCGCCACCGCTGACACAGGATGTGAACGCT 1381
1421 GTGAATCT 1480
1382 GTGAATCT 1441
1481 CAGAGTGGCAGATGATGCTGAGCTATGCACTGCACTGCTCTCTCTCTCTCTCTCTCT 1540
1442 CAGAGTGGCAGATGATGCTGAGCTATGCACTGCACTGCTCTCTCTCTCTCTCTCTCT 1501
1541 ACCATGATTTTTCAGAGTGTGTCTACTGAGCGCTGGGCTTACATCTGAGTTGGCCGA 1600
1502 AGCGTATTTTTCAGAGTGTGTCTACTGAGCGCTGGGCTTACATCTGAGTTGGCCGA 1561
1601 TTTGATATTTTTCAGAGTGTGTCTACTGAGCGCTGGGCTTACATCTGAGTTGGCCGA 1660
1562 TTTGATATTTTTCAGAGTGTGTCTACTGAGCGCTGGGCTTACATCTGAGTTGGCCGA 1621
1661 GTGCTACTTGTATCTCTCAAAATCAATATAGTCAAGAGTATGCTGCTGCTGCTGCTG 1720
1622 GTGCTACTTGTATCTCTCAAAATCAATATAGTCAAGAGTATGCTGCTGCTGCTGCTG 1681
1721 ATTGCGGGGCTCTCTTGTGATCTGATCTCTGAGAAACATCAAGAAATGCCATCTCT 1780

Db 1682 ATTCTGGGTTGCTCATGATGATCTGATTTATCAGAAACATAGAAAGATGCCATTCT 1741
OY 1781 TTTAAATCATCATGATTTATTTACATCCAAAATATTCAGATTCCTGTACTCAAT 1840
Db 1742 TTTAAATCATCTGGGTTACTTACCTTCCAAAAGTCTGTGATTCCTGTGATCAAT 1801
OY 1841 GAGTCTACGAGCTGAATTCATCTGTGGCAGCTCAATGTTTCTGTACAACTATATCA 1900
Db 1802 GAGTCTATGAGCTGCAATCTCATCTGTGGTCCAACTCTGTGTCGCAAAATACCA 1861
OY 1901 ATGTGCTCTACATCAGAAATTCATTCATTCAGAAACCTGCCAGGTCACATCT 1960
Db 1862 ATGTGCTCATGACCAAGGATTCATTCATTCAGAAACCTGCCAGGTCACATCT 1921
OY 1961 AGATTCACATGAACTTCTGATTTCTATTCATTTATTCAGCTCTGTCTATCTAGGA 2020
Db 1922 AGATTCACGACAACTCTGATCTGTGATCTGTTACCTCCGACCTGTCTATCTGATG 1981
OY 2021 ATAGTCTCTTCAAAATTAAGGATCATCTCATTCAGAGTGTGAAAGCCATGCTGGGA 2080
Db 1982 ATGTGCTCTTAAAGTCCGGGACTCACTGATTCAGATAGTAGATGAGCAGGACAGA 2041
OY 2081 AATGGAAGTG 2091
Db 2042 AAGGTTAATG 2052

RESULT 13
AX685729 1959 bp DNA linear PAT 29-MAR-2003
LOCUS AX685729
DEFINITION Sequence 1 from Patent WO02081691.
ACCESSION AX685729
VERSION AX685729.1 GI:29371738
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1
AUTHORS Hobbs, H.H., Shan, B., Barnes, R., and Tian, H.
TITLE Abcg5 and abcg8: compositions and methods of use
JOURNAL Patent: WO 02081691-A 1 17-OCT-2002;
Tularik Inc. (US); BOARD OF REGENTS UNIVERSITY OF TEXAS SYSTEM
(US)

FEATURES
Source Location/Qualifiers
1..1959
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
1..1959
/note="unnamed protein product; ABCG5 (mAbCG5)"
/codon_start=1
/protein_id="CAD86570.1"
/db_xref="GI:29371739"
/translation="MSELPLSPGANGPHINRGSLSLQGSVTGTEARHSGLVHV
SYSNRVGFPMNIRKSCQQRMDRLIDVSLYIESGIMCILLSSGSKTLLDAISG
RLRTGTLBEEVFNELSRDQFODCFSVLQSDVLSLIVETLRVYALALCS
SADFYNNKVAEYVETLSLHVADIMGSIYNGFSGSERRRVSLAALLDPKMMID
EPTGLDCAANOIVLLAEARDRIVITIHOPRESELHPDKRIILYGEIVFGC
TPREMLGEFNNCGPCPEHNSNPDEYMDLVSDPOSRELERIYKKROMLECFKESD
IYKILIEIRARYIKTLPMVPEPKTRDPRMEKGLVLLRVRNLRNKRQAYIMRLV
ONLMGLFLFLIRYONNTLKGAVDRVGLTYQVCAATPYTGLNANVLEPMLRAVS
DOESODELYHKQWMLAYVLAIVLPEFVIAIVIFSSVCWYTLGLYEVARGYSALL
APHLIGELFLVLEGIYONPNIVNSIVALLISGLISGFIIRINIOEMPIPLKIGYF
TPQYCCCEILVNEFYGLNFTCGSNTSMNHMPCAITQGVOFIEKTCPGATSHFTAN
FLIYGFIPRLVILGIYIFKVRDYLLSR"

BASE COUNT 468 a 506 c 495 g 490 t
ORIGIN

Query Match 58.4%; Score 1365.4; DB 6; Length 1959;
Best Local Similarity 81.4%; Pred. No. 1e-286;
Matches 1595; Conservative 0; Mismatches 361; Indels 3; Gaps 1;

OY 107 ATGGGTGACCTCTCATCTTTTAAACCCCGGAGGCTCATGCTCTCCAAAGTAACAGAGGC 166
Db 1 ATGGGTGACCTCTCTTTCTGATGTCACAGAGGAGCAGAGGCGCTCACATCAACAGAGG 60
OY 167 TCCAGAGCTCCCTGGAGAGGGGCTCTCCACACCGCCCGGAGCT---CACAGCTGGGC 223
Db 61 TCTGAGACTCCCTGGAGCAGAGGTTGGTCCACAGGCGCACAGAGGCTCGGACAGCTTAGGT 120
OY 224 ATCCCTCAATGCTCTTACAGGCTCAGCAGCCGCGTGGGCTGGTGGATCATCATCT 283
Db 121 GTCTGCAATGTCTCTACAGGCTCAGCAGCCGTGTGGGCTTGGTGGATCATCAATCA 180
OY 284 TCCGCGCAGAGTGGACAGGACAGATTCCTCAAAAGTGTCTCTGTACAGTGGAGAGCGGG 343
Db 181 TGCCAGCAGAGATGGAGACAGCAATCTCAAAAGTGTCTCTGTACAGTGGAGAGTGGC 240
OY 344 CAGATCATGTGCTATCTAGAGAGCTCAGGCTCGGGGAAACACAGCTGCTGAGAGCCATG 403
Db 241 CAGATTAATGCTCATCTTAAAGCAGCTCAGAGGCTCAAGGAAACACAGCTGCTGAGAGCCATC 300
OY 404 TCCGAGAGCTGGGAGCGCGGAGACCTCTCTGGGGAGAGGTGTGTGAACAGCGCGGCG 463
Db 301 TCCGAGAGCTGGGAGCGCGGAGACCTCTCTGGGGAGAGGTGTGTGAATGAGCTGCGAG 360
OY 464 CTGCGCGCGGAGAGTTCACAGAGCTGCTCTCTACGCTCGACAGAGGACACCTGCTG 523
Db 361 CTGCGAGAGGAGCAGGTTCCAGAGACTGCTCTCTCTACAGCTCTGCGAGAGAGCTTTTCTG 420
OY 524 AGCAGCTCAGCTGCGGAGAGCGCTGACATACACCGCGCTGTGGCCATCGCGCGGC 583
Db 421 AGCAGCTCAGCTGCGGAGAGCGCTGACATACACAGAGAGTGTGGCCCTGTGCGCGAGC 480
OY 584 AATCCGCGCTCTTCCAGAAAGTGGAGGCGCTCATGAGAGCTGAGCTAGCAT 643
Db 481 TCCGCGAGCTCTTCAACAAAGAGTGAAGGACATCAAGAGAGCTGAGCTAGGCAC 540
OY 644 GTGCGAGACGCACTGATTTGGCACTACAGCTTGGGGGCAATTTCCAGGGTGAGCGGCG 703
Db 541 GTGCGAGACCAATGATTTGGAGCTATAATTTGGGGGAATTTCCAGTGGGAGCGGCGC 600
OY 704 CGGCTTCATCTGAGCGGAGCGCTCTCAGAGATCTTAAGTCACTGCTTTGATGAGCCA 763
Db 601 CGAGTTTCATCTGAGCGGAGCGCTCTCAGAGATCTTAAGTCACTGATGAGAGCCA 660
OY 764 ACCAGAGCGCTGCTCATGACTGCTAATCAGATTTGTGCTCGCGGAGAGCTGGCT 823
Db 661 ACCAGAGACTGCTCATGACTGCTAATCAGATTTGTGCTCGCGGAGAGCTGGCT 720
OY 824 CGCAGGACCGAATTTGATTTCTCACCATTCCAGCGCCGTTCTGAGCTTTTTCAGCTC 883
Db 721 CGCAGGACCGAATTTGATTTCTCACCATTCCAGCGCTGCTGAGCTTTTCAGCTC 780
OY 884 TTTACCAAAATTTGCTGCTGAGCTTGGAGAGCTGATTTTCTGTGCGCAGCGGAA 943
Db 781 TTTACCAAAATTTGCTGCTGAGCTTGGAGAGCTGATTTTCTGTGCGCAGCGGAA 840
OY 944 ATGCTTATTTCTTCAATGCTGCGGCTTACCTTTGCTGCTAATCAAAACCTTTGAC 1003
Db 841 ATGCTTATTTCTTCAATGCTGCGGCTTACCTTTGCTGCTAATCAAAACCTTTGAC 900
OY 1004 TTTATATGAGCTGAGCTGAGTGTATCCCAAGAGAGGAGGAGGAAATGAAGAACTCC 1063
Db 901 TTTATATGAGCTGAGTGTATCCCAAGAGAGGAGGAGGAAATGAAGAACTCC 960
OY 1064 AAGAGATCCAGATGATGATGCTGCTTACAGAAATCAGCAATTTGCTATTAAGCTTTG 1123
Db 961 AAGAGATCCAGATGATGATGCTGCTTACAGAAATCAGCAATTTGCTATTAAGCTTTG 1020
OY 1124 AAGATATTTGAAGATGAAGACACTGAAGAACTGATGCTTCTTCAAAACCAAA 1183
Db 1021 GAGAGACTTGAAGAGAGAGATGCTGAAACCTTACCTGCTTCTTCAAAACCAAA 1080

1184 GATTCTCTGAGAGTTTCTCTAACTGGGCTTCTCTAGAGAGAGTCAAGAACTTG 1243
1081 GATCCCTGCGGAGTGTGGCAAGCTGTGCTGCTGAGCGAGGATACAAAGAACTTA 1140
1244 GTGAGAAATAGCTGAGTGAATAGCGTCCCTTCAAGATCTGATGCTGCTGCTG 1303
1141 ATGAGAAATAGCAGAGCACTGATATGCTTCTGCTCAAGATCTGATGCTGCTGCT 1200
1304 CTCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1363
1201 CTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
1364 GTAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1423
1261 GTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
1424 AATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1483
1321 AATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
1484 AAGTGGCAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1543
1381 AAGTGGCAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440
1544 ATGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1603
1441 GTCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
1604 GATATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1663
1501 GATATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560
1664 CTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1723
1561 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
1724 GCGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1783
1621 TCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
1784 AAAATCATCATGTTATTTACATTCGAAATATGCAATGCAATGCTGCTGCTGCTG 1843
1681 AAAATCATCATGTTATTTACATTCGAAATATGCAATGCAATGCTGCTGCTGCTG 1740
1844 TTCTACGAGCTGAAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1903
1741 TTTTACGGGCTGAACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800
1904 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1963
1801 TGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
1964 TTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2023
1861 TTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
2024 GTTGTCTTCAAAATAGGATCATCTATAGCAGGTAG 2062
1921 GTGATTTTAAAGTCAGGAGCTACCTGATTTAGCAGATAG 1959

RESULT 14
AX456526
LOCUS AX456526 2035 bp DNA Linear PAT 06-JUL-2002
DEFINITION Sequence 48 from Patent WO0227016.
ACCESSION AX456526
VERSION AX456526.1 GI:21715414
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1
AUTHORS Patel,S.B. and Dean,M.

TITLE Gene involved in dietary sterol absorption and excretion and uses therefor
JOURNAL Patent: WO 0227016-A 48 04-APR-2002;
THE DEPARTMENT OF HEALTH AND HUMAN SERVICES (US) ; Patel, Shalendra B. (US) ; Dean, Michael (US)
FEATURES
source 1. 2035
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="Primer"
BASE COUNT 481 a 533 c 537 g 484 t
ORIGIN
Query Match 58.2%; Score 1363; DB 6; Length 2035;
Best Local Similarity 80.6%; Pred. No. 3.5e-286;
Matches 1607; Conservative 0; Mismatches 385; Indels 3; Gaps 1;
100 GTTGGCAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 159
1 GCTGGCAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60
160 CAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 216
61 CAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
217 CCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 276
121 CTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
277 CACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 336
181 CAATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
337 GAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 396
241 GAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
397 GCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 456
301 GCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
457 GCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 516
361 CTGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
517 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 576
421 CTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
577 CCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 636
481 CAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
637 GAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 696
541 GAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
697 GCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 756
601 GCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
757 TGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 816
661 CGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 720
817 ACTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 876
721 GCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
877 TCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 936
781 CCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840

QY 2017 AGGAATGTTCTTTTCCAAAATAAGGATCATCTATTAGCAGGAGTGAAGACCATGCT 2076
DB 1921 GCGGCTGGTGGCTTTTAAAGTCCGGGACTACCTATTAGCAGATAGGTAAAGATGGCAGGC 1980
QY 2077 GCGAAAATGGAAGTG 2091
DB 1981 AGGAAGGGTTAAATG 1995

and uses

2100 00000

Gaps 1;

AGAGGC 166
|||||

AGAGGG 60

TTGGGC 223

11
TAGGT 120

CATCT 283

111
180

00T WOLFF

GC666 343
111

GTGGC 240

CCATG 403
1111

CCATC 300

GGGCG 463

1111
GCCGAG 360

523

[illegible]

REC-76 420

583 5CGGC 1111

3CAGC 480

SCCAT 643

—

Mon Jul '28 09:43:27 2003

us-09-989-981a-5.rge

Page 21

Db	481	TCGCGGACCTTCTACACAAAGAAAGTAAAGGACGATCATACAGAGCTGAGCCTGAGCCAC	540
QY	644	GTGGCAGACCCGACGTGATTGGCACTACAGCTTGGGGGGGCAATTTCCACGGGTGAGCGGCGC	703
Db	541	GTGGGGGACCAATATGATTTGGCAGCTGTAAATTTTGGGGGAATTTCCAGTGGCGACGGCGCC	600
QY	704	GGGGTCTCCATGGCAGCCGAGCTGCTCCAGGATCCTAAGGTATGCTGTTGATGAGCCA	763
Db	601	CGAGTTTCCATGCGAGCCCACTCCCTTCAGGAGCCCAAGGTATGATGCTATGATAGCCA	660
QY	764	ACCAGAGGCGCTGAGCTGATGACTGCTATGATTTGGTGTCTCTCTGGTGGAACTGGCT	823
Db	661	ACCAGAGGACTGAGCTGATGACTGCTATGATTTGGTGTCTCTCTGGTGGAACTGGCT	720
QY	824	CGCAGGAACCGAATGTGTGGTTCACCATTCACACAGCCCGCTTGTGAGCTTTTTCAGCTC	883
Db	721	CGCAGGAGCCGAATGTGTGATGTGACACATTCACACAGCCCGCTTGTGAGCTTTTTCACAC	780
QY	884	TTTGGCAAAATTTGCCATCTCTGAGCTTCGGAGAGCTGATTTTCTGTGGCACGCCAGGGAA	943
Db	781	TTTCGACAAAATTTGCCATCTCTGACTTACGAGAGATGTGTGTCTGTGGCACCCCAAGAGAG	840
QY	944	ATGCTGTGATTTTCTCAATGACTGCGGTTAACCCCTTGCCGACATTCACAAOCCCTTTTGC	1000
Db	841	ATGCTGTGCTTTTCAATTAAGCTGTGTGTTAACCCCTTGCCGACATTCACAAOCCCTTTTAT	900
QY	1004	TTCTATATGACCTTACGCTCAGTGTGATACCCAAAGCAGAAAGGGAAATAGAAATCTCC	1063
Db	901	TTTTCATGAGCTTATGATACATGATGTGACACCCAAAGCAGAGCGGGGAATATAGAAATCTAC	960
QY	1064	AAGAGATGCCAGATATAGAAATCTCTCCCTACAAAGAAATACGAATTTGTATATAAATCTTG	1123
Db	961	AAGCAGATACAGATCTGGAATGTCTCTTCAAGAAATGTGACATCTACACAAATTTCTG	1020
QY	1124	AAGATATTTGAAAGAAATGAACACCTGAAACGTTACCAAGTGTCTCTTCAAAACAAA	1183
Db	1021	GAGAACATTTGAAAGAGCAGATACCTCTGAAACCTTACCCAGGTTCTCTTCAAAACAAA	1080
QY	1184	GATTCCTCGGAGTTTCTCTAAATCTGGGTGTCTCTGAGAGAGTACAAAGAACTTG	1243
Db	1081	GATTCCTCGGAGTTTCTCGGCAAGTTTGCTGTCTGTGCTGTGAGGCGAGTACAAAGAACTTA	1140
QY	1244	GTGAGAAATAAAGCTGGCAGTATTAACGCGTCTCTTCAGAAATCTGATCATGGTTTGTTC	1303
Db	1141	ATGAGGAATAAAGCAGGAGATTAATATGCGTCTCTGTTCAGAAATCTGATCATGGGCTTTC	1200
QY	1304	CTCCCTTTCTGCTCTGCGGGGTCGGAAGCAATGTGCTAAAGGGTGTCTATCCAGAGCCG	1363
Db	1201	CTCATTTTCTACCTTCTCCGCGCTCAGAACAAACGCTTAAGGGGCGCTGTCCAGAGCCG	1260
QY	1364	GTAGGTCTCTTTACCAAGTTTGTGGGCGCCACCCGTTACACAGGATGCTAAGGCTGTG	1423
Db	1261	GTGGGGCTGTCTTATACGCTTGTGGGTGCCACCCCATATCACCGGATCTTAATGCTGTG	1320
QY	1424	AATCTGTTCCCGTGTCTGACGCTGTACGAGCAAGAGAGTACAGAGCGCTTACACG	1483
Db	1321	AATCTGTTCCCAATGCTGAGAGCCGTCAGCAGCAGAGAGTACAGAGTGGCTGTATCAT	1380
QY	1484	AAGTGGCAGATGATGAGGCGCTTATGCACTGTGACGCTCTCCCTTCAGCGTTGTGGCAC	1543
Db	1381	AAGTGGCAGATGATGAGGCGCTTATGCACTGTGACGCTCTCCCTTCAGCGTATGACGACG	1440
QY	1544	ATGATTTTTCACAGATGTGTCTACTGTGACGCTGGGCTTACATCTCTGAGAGTTGCCGATTT	1603
Db	1441	GTCTATTTTTCACAGATGTGTGTATTATGGACGTGTGGGCTGTATCTCTGAAGTTGCCAGATTT	1500
QY	1604	GGATATTTTTCGTGCTGTCTTGTGGCCGCCACTTAATGGTGAATTTCTAATCTTGTG	1663
Db	1501	GGATATTTTTCGTGCTGTCTTGTGGCCGCCACTTAATGGAGAAATTTCTAATCTTGTG	1560
QY	1664	CTACTTGTATTCGTCGCAAAATCCAAATATATGTCAACAGTATAGAGCGCTGTGCTGCATTT	1723
Db	1561	CTGCTGTGTATGTCGCAAAACCTTAATATGTCTCAACAGTATATAGTGTGTGCTGTGATGATC	1620

OY	1724	GCGGGGGGGCGTGTGGATCTGGATTTCTCAGAAACATACAGAAATGCCATTCCTTTT	1783
Db	1621	CTGTGGGCTGCTATTTAGTATCTGGATTTATTCAGAAACATACAGAAATGCCATTCCTTTA	1680
OY	1784	AAATTCATCACTTATTTTACATTTCCAAAATATTTGCAGTGAATCTTGTAGTCAAATGAG	1843
Db	1681	AAATTCCTGGGTTTATTTTACATTTCCAAAATATCTGTTGTGAGATTTCTGTGTCAATGAG	1740
OY	1844	TTTCAGGACATTTTCACATTCCTTGGCGCAGCTCAATGTTTCTGTGACAACTAATCCAAATG	1903
Db	1741	TTTTCAGGCTCGAACTTCATCTTGGTGGAGATCCCAACACCTCTTAATGCAAAATACCCCATG	1800
OY	1904	TGTGCTCTTCACTCAAGGAATTCATTTGATGAGAAAACCTGCCAGGTGCACAACTAGAGA	1963
Db	1801	TGCCTCCATTCACCCCAAGGGGTCCACTTCATTCGAGAAAACCTGCCAGGTGCTCAATCCAGA	1860
OY	1964	TTTCACATGAACCTTTCGATTTTGTATTCATTTATTCAGAGCTCTTGTCAATCTTAG	2018
Db	1861	TTTCAGGCAAACTTCTCATCTTATTAATTAATGGTTATTCAGAGCTCTTGTGTCAATCTTAG	1915

Search completed: July 27, 2003, 05:52:29
Job time : 8322.43 secs

us-09-989-981a-5.rni

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: JULY 27, 2003, 05:52:44 ; Search time 147.622 Seconds
(without a) : 147.622

6996.481 Million cell updates/sec

Sequence: 1 gtcagtgtagcagcagcag...atatcatataacgtatcttc 3360

Scoring table: IDENTITY_NUC

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Maximum Match 1008
Listing first 45 summaries

```
Database : Issued_Patents_NA:*
```

```

1: /cgn2_6/ptodata/1/lna/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/lna/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/lna/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/lna/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/lna/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/lna/PCTUS_COMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	98	4.1	2418	4	US-09-245-808-2	Sequence 2, Appl
2	96.6	235	3	US-09-172-108-8	Sequence 8, Appl	
3	83	3.5	3376	4	US-09-620-112D-918	Sequence 918, Appl
4	57	2.4	1668	4	US-09-252-991A-13705	Sequence 13705, A
5	56.8	2.4	4403765	3	US-09-103-840A-2	Sequence 2, Appl
6	56.8	2.4	4411529	3	US-09-103-840A-1	Sequence 1, Appl
7	56.2	2.4	972	4	US-09-252-991A-9848	Sequence 1, Appl
8	56.2	2.4	1713	4	US-09-252-991A-9760	Sequence 9848, A
9	56.2	2.4	2805	4	US-09-252-991A-10208	Sequence 9760, A
C 10	55.6	2.4	1509	4	US-09-252-991A-13436	Sequence 10208, A
11	54.2	2.3	840	4	US-09-252-991A-15851	Sequence 13436, A
12	54	2.3	2175	4	US-09-252-991A-13557	Sequence 15851, A
13	54	2.3	2292	4	US-09-252-991A-10395	Sequence 13557, A
14	53.4	2.3	993	4	US-09-252-991A-10934	Sequence 10395, A
15	53.4	2.3	1335	4	US-09-252-991A-10934	Sequence 10934, A
C 16	53.4	2.3	2178	4	US-09-252-991A-11254	Sequence 10934, A
C 17	53	2.3	4403765	3	US-09-103-840A-2	Sequence 11254, A
C 18	53	2.3	4411529	3	US-09-103-840A-1	Sequence 2, Appl
C 19	51.6	2.2	765	4	US-09-252-991A-1138	Sequence 1, Appl
20	51.6	2.2	1476	4	US-09-252-991A-8825	Sequence 3138, Ap
21	51	2.2	945	4	US-09-252-991A-8027	Sequence 8825, Ap
C 22	51	2.2	1173	4	US-09-252-991A-8287	Sequence 8027, Ap
23	50.6	2.2	765	4	US-09-252-991A-16348	Sequence 8287, Ap
24	50.6	2.2	846	4	US-09-252-991A-16477	Sequence 16348, A
C 25	50.6	2.2	855	4	US-09-252-991A-16015	Sequence 16477, A
C 26	50.6	2.2	867	4	US-09-252-991A-15936	Sequence 16015, A
27	50.6	2.2	2427	4	US-09-252-991A-16255	Sequence 15936, A
						Sequence 16255, A

28	50	2.1	744	4	US-09-252-991A-13301	Sequence 13301, <i>h</i>
29	50	2.1	783	4	US-09-252-991A-12981	Sequence 12981, <i>h</i>
30	50	2.1	786	4	US-09-252-991A-12498	Sequence 12498, <i>h</i>
31	50	2.1	21706	4	US-08-961-527-36	Sequence 36, <i>Appl</i>
32	49.4	2.1	1239	4	US-09-252-991A-8289	Sequence 8289, <i>Ap</i>
33	49	2.1	909	4	US-09-252-991A-5320	Sequence 5329, <i>Ap</i>
34	49	2.1	1164	4	US-09-252-991A-5240	Sequence 5240, <i>Ap</i>
35	49	2.1	1620	4	US-09-252-991A-5364	Sequence 5364, <i>Ap</i>
36	49	2.1	2067	4	US-09-252-991A-5209	Sequence 5209, <i>Ap</i>
37	49	2.1	2347	4	US-09-252-991A-5294	Sequence 5294, <i>Ap</i>
38	49	2.1	5975	1	US-08-404-0548-1	Sequence 1, <i>Appl</i>
39	49	2.1	5975	1	US-08-414-0838-1	Sequence 1, <i>Appl</i>
40	49	2.1	5975	1	US-08-436-6758-1	Sequence 1, <i>Appl</i>
41	49	2.1	5975	1	US-08-435-6758-1	Sequence 3, <i>Appl</i>
42	49	2.1	5975	3	US-08-884-559-1	Sequence 1, <i>Appl</i>
43	48.8	2.1	738	4	US-09-252-991A-4086	Sequence 4086, <i>Ap</i>
44	48.8	2.1	837	4	US-09-252-991A-4070	Sequence 4070, <i>Ap</i>
45	48.8	2.1	966	4	US-09-252-991A-4119	Sequence 4119, <i>Ap</i>

ALIGNMENTS

RESULT 1

US-09-245-808-2

; sequence 2, Application No. 6313277;
; Patent No. 6313277

```

;; GENERAL INFORMATION:
;;
;; APPLICANT: Doyle, T

```

APPLICANT: Abruazzo, Lynne V.

TITLE OF INVENTION: Breast

```

; FILE REFERENCE: ROSS UMB conversion
; TITLE OF INVENTION: encodes it

```

CURRENT APPLICATION NUMBER: US/09/245,808
CURRENT FILING DATE: 1999-03-05

EARLIER APPLICATION NUMBER: 60/073763
EARLIER FILING DATE: 1999-03-02

NUMBER OF SEQ ID NOS: 7

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2

```

```

; LENGTH: 2418
; TYPE: DNA

```

ORGANISM: Human MCF-7/Adrvp cells

2
0
0
0
7
3
0
0
0

Query Match	4.28;	Score 98;	DB 4;	Length 2418;
Best Local Similarity	50.28;	Pred NO	5	60.17.

Matches	242;	Conservative	0;	Mismatches	240;	Indels	0;	Gaps	0
---------	------	--------------	----	------------	------	--------	----	------	---

617 GTCATGGCAGACTGAGTCTGAGCCATGTCGACAGACCGACTGATTGGCAACTACAGCTTTC 676

728 GTCATTCAGAGTTAGCTGGATAAAGTGGCAGACTCCAGGCTTCCAACTCCCTTCTG

[illegible]

736
788

CGGCTGGTGTGTTTGATGAGGACCTAGTATAGGAATGCAGCTTATCACTGAT 847

3/ CCGAGGTCATGCTGTTTGATGACCAACACAGGCGTGCATGCATGACTGCTAATCAG 796

848 CCTTCATCTTGTCTTGATGACCTACACTGGCTTAGACTGCAAGCACAGCAATGCT 907

797 ATTCTCGTCTCTGTGGAACAGGCTCGCAGGAACCGAATTCGTTCTCACCATTTCAC 856

908 GTCTTTTGGCTCTGAAGCATGCTTAAGCAGGAGCAACATCATCTTCTCCATTCAT 957

[illegible]

968 CAGGCTTCGATATTTCCACATCCTCAACCTCGAGGCGTGGAGAG **976**

017 CTTCTTGGCCCTCAGGAGA 1027

27. CCGGATTCGACGCTGCGTTACCT 976

Query Match	2.48;	Score 57;	DB 4;	Length 1668;
Best Local Similarity	45.78;	Pred. No. 1.1e-05;		
Matches 238;	Conservative	0;	Mismatches 280;	Indels 3.

RESULT 5
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6304230

```

1 GENERAL INFORMATION:
2 APPLICANT: FLEISCHMAN, Robert D.
3 APPLICANT: WHITE, Owen R.
4 APPLICANT: FRASER, Claire M.
5 APPLICANT: VENTER, John C.
6 TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
7 TUBERCULOSIS

```

Query Match	2.4%	Score 56.8;	DB 3;	Length 4403765;
Best Local Similarity	46.0%;	Pred. No. 0.0009;		
Matches 273;	Conservative	0;	Mismatches 312;	Indels 9;
				Gaps 2;

RESULT 6
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6304330

APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103.840A

RESULT 7
 US-09-252-991A-9848
 Sequence 9848 Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 9848
 LENGTH: 972

QY 311 CTGAAGATGTCCTCTGTCCTGTCACGTGAGAGAGAGGGGCGACATCTATGTGCATCTCTAGAACCTCA 370

Db 88 CTGAACGGCGCTATTCGTGTCAGTGCAGAGGCCGGGAAGACCTCCGCTGGTCGGCGAGTCC 147

QY 371 GGCTCCGGGAAAAACAACGCTGCTGCAGACGCATCTCCGGAGAGGCGTGGGGGCGCGGGAGACC 430

Db 148 GCGTCGGGCAAGTCGACACCTCGGGCGCGCGCTTCACCTCATCTGACAGAACCCACCTCCGCG 207

QY 431 TTCTCTGGGGAGGTGTATGTGAAAGGGCCGGGCGCTCGCGCGGGAGCAGTTCCAGGACTCG 490

Db 208 TCGCTGAAAATCCCGCGGCGAGGAGTCAAGGGCGCCAGCAAGGAGACAGCGCCGCA ---G 264

QY 491 TTCTCTACGTCCTGCGAGAGCGACACCTCTGTAGACAGCTTACACCTGCGCGGAGACGCTG 550

Db 265 TTTCGCGCGCGACCTGCAGATGGTCTTCCAGAACCCCTACGCGCTCACTCAATCCGCGACAG 324

QY 551 CACTACACCGCGCTGCTGGCCATCCGCGCGGCAATCCGGCTCTCTCCAGAAAGAGTg 610

Db 325 AAGATCGGGAGACCAAGTTGGCGCGAGCGGCTGTGTATATACACCGCGCTGTCCGCGGAGGAA 384

QY 611 GAGCGCTATGGCAAGCTGAGTCTGAGCCATGTGGCAGACCGACTGATTGGCAACTAC 670

Db 385 CGCGCGAAAAAGGTCACGAGATGAATGACCCAGTGGCGCTGGCGCGCGAGCATTTACAG 444

QY 671 AGCTTGGGGGGGCTTTCACAGGGTG--AGCGGCGCGGGGTCTCCATATCCGACCCAGCTG 727

Db 445 CGGTACCCGCACATGTCTTCGCGCGGCCACGGCCAGCGCATGCGCTCGCGCGGGGAGT 504

QY 728 CTCACAGATCTTAAGTTCATGTCTGTTTGAATGAGCCAAACCAACGAGGCTGTGACTGCATGACT 787

Db 505 ATGCTGTCAACCCAAAGTGTCTGGTGGGAGAGACGCACTCGGCGCTGCAAGTGTGATG 564

QY 788 GCTAATCAGATTGTGCTCTCTCGTGGGGAACGTGGCTCGGAGAACCCGAATTGTGTTCTC 847

Db 565 CAGGCCCAAGTACTGAACCTTTTCATGAGCACTGCACAGCAAGTTCGCGACCGCTACAGTG 624

QY 848 ACCATTACCAAGCCCGCTTCTGAGCTTTTTCAGCTCTTTCAGCTAAATTTGCCATCTCGAGC 907

Db 625 TTTCATCTGCACAACTGGCGGGTGGTCCGCCACAGTGTCCGAGAGAGACTCTCTGTGATGTAC 684

QY 908 TTCGG 912

Db 685 CTCGG 689

RESULT 8
 US-09-252-991A-9760
 : Sequence 9760, Application US/09252991A
 : Patent No. 6551795
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Marc J. Rubenfield et al.
 :
 : TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 :
 : TITLE OF INVENTION: AERuginosa FOR DIAGNOSTICS AND THERAPEUTICS
 :
 : FILE REFERENCE: 107196.136
 :
 : CURRENT APPLICATION NUMBER: US/09/252,991A
 :
 : CURRENT FILING DATE: 1999-02-18
 :
 : PRIOR APPLICATION NUMBER: US 60/074,788
 :
 : PRIOR FILING DATE: 1998-02-18
 :
 : PRIOR APPLICATION NUMBER: US 60/094,190
 :
 : PRIOR FILING DATE: 1998-07-27
 :
 : NUMBER OF SEQ ID NOS: 33142
 :
 : SEQ ID NO 9760
 :
 : LENGTH: 1713
 :
 : TYPE: DNA
 :
 : ORGANISM: Pseudomonas aeruginosa

us-09-989-981a-5.rni

US-09-252-991A-9760

Query Match	2.48;	Score 56.2;	DB 4;	Length 1713;
Best Local Similarity	45.68;	Pred. No. 1.8e-05;		
Matches 276;	Conservative	0;	Mismatches 323;	Indels 6;
			Gaps 2;	

[illegible]

RESULT 9
 US-09-252-991A-10208/C
 : Sequence 10208, Application US/09252991A
 : Patent No. 6551795
 : GENERAL INFORMATION:
 : APPLICANT: Marc J. Rubenfield et al.
 : TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 : TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 : FILE REFERENCE: 107196.136
 : CURRENT APPLICATION NUMBER: US/09/252,991A
 : CURRENT FILING DATE: 1999-02-18
 : PRIOR APPLICATION NUMBER: US 60/074,788
 : PRIOR FILING DATE: 1998-02-18
 : PRIOR APPLICATION NUMBER: US 60/094,190
 : PRIOR FILING DATE: 1998-07-27
 : NUMBER OF SEQ ID NOS: 33142
 : SEQ ID NO 10208
 : LENGTH: 2805
 : TYPE: DNA
 : ORGANISM: Pseudomonas aeruginosa
 : US-09-252-991A-10208

Query Match	2.48;	Score 56.2;	DB 4;	Length 2805;
Best Local Similarity	45.68;	Pred. No. 2.4e-05;		
Matches 276; Conservative	0;	Mismatches 323;	Indels 6;	Gaps 2;

QY 311 CTGAAGATGTCCTCTTGTACGAGAGACGGGAGCATATGTGATCTCTTAGAAGCTCA 370
 Db 1629 CTGAAGGGGTATCGTTCCAGTGTGAAGCCGGCAAGACCTTCGCGGTGTGGCGAATCC 157/0
 QY 371 GGCTCGGGAAAAACCAACGCTGCTGACAGCCATGTCCTGGGAGGCTGGGGGGGGGGAGCC 430
 Db 1569 GGGTGGGGCAAGTGCAGCCCTGGCGGGGGGCCCTGACACCTGATTCAGAGAAACCACTCCGGC 151/0
 QY 431 TTCTCTGGGGAGGTGATGTGAACGGCGGGCGCTGGCGCGGAGCAATTCAGAGATGTC 490
 Db 1509 TGGCTGAATAATCGCGGGGAGGAGGTCAAGGGCGCCACAGAGAACACGCGCGGGA---G 145/5
 QY 491 TTCTCTACGTCCTCTCAGAGCAGACCCGTCGAGACACCTCACCGTCCGGAGACGCTG 550
 Db 1452 TTGGCGCGGAGCTGCAGATGTCTCTTCAGAAACCCCTACGCTCTGCTTAATCCCGACAG 1393/3
 QY 551 CACTACACCGCGCTGCTGGCCATCCCGCGGCAATCCCGGCTCTTCACAGAAAGGTG 610
 Db 1392 AAGATCGCGACACGATTGGCGAGCGCGTGTGATCAACACCGCGCTGTCCGGGAGAA 1333/3
 QY 611 GAGGCGGTCATGGCGAGACTGATGTCAGCCATGTGGCAGACGATGATTGGCACTAC 670
 Db 1332 CGCGCGGAAAGGTCCACGATGATATGCGCCAGGTGCGGCTGGCGCGGAGCAATTACAG 1273
 QY 671 AGCTTGGGGGCAATTTCCACGGGTG---AGCGGGCGGGGTCCTCCATTCGACGCCAGCTG 727
 Db 1272 CGCTACCGGCACATGTTCTCCGGGGCCAGCGCCAGCGCATTCGCTCGCGCGGCGCAT 1213
 QY 728 CTCAGAGTCTTAAAGTATGTCGTTTGTATGAGCAACCAAGCGCTGAGACTGCATGACT 787
 Db 1212 ATGTGTGCAACCAAGTGTGTGGCGGAGAGAGCACTCGGGCCCTCGACGCTGTGATC 1153
 QY 788 GCTATATGATTTGCTGTCCTCTGTGTGGAAGTGGCTGCGAGAACCGAATTTGTGTTTC 847
 Db 1152 CAGGCGCAAGTACTGAACCTTTTATATGAGCTCAGAGCAGTTCGCGACCGCCTACGT 1093
 QY 848 ACGATTACCAAGCCCCGTTTGAAGCTTTTACAGCTCTTTACAAAATTTGCATTCAGAC 907
 Db 1092 TTTCAATTCGCAACACGTGGGGGTGTGTCCGCACGTGCGCAGACAGCTCTCGTGTGATAC 1033
 QY 908 TTGCG 912
 Db 1032 CTGCG 1028

```

RESULT 10
US-09-252-991A-13436/C
: Sequence 13436, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252.991A
: PRIORITY FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074.788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094.190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 13436
: LENGTH: 1509
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
: US-09-252-991A-13436

```

Query Match	2.48;	Score 55.6;	DB 4;	Length 1509
Best Local Similarity	46.08;	Pred. No. 2.5e-05;		

	Query Match	2.3%	Score 54.2	DB 4	Length 840
	Best Local Similarity	46.4%	Pred. NO. 4.e-05		
	Matches	253	Conservative	0	Mismatches 283
					Indels 9
					Gaps 2
QY	318	ATGTCCTCTTACTGGAGAGCGGCGCATCATCTGCATCTTACGAAAGCTTACGGCTCG	377		
Db	107	ATTCCTCTGCTGGCCATCCCGAGGGCTTGTTACGTATGATCTGCGGCCCAACGGCTTCG	166		
QY	378	GGAAACACAGCTGCTGGAGCCCATGTCTGGGAGGCTGGGGCGCGCGGGAGCTTCTTCGG	437		
Db	167	GCAAGTCACACCTCTGGCGGCAATTGTCGCCCTCTGTTGGCCCGGCGCCAGGGCCGGGTGG	226		

RESULT 12
 US-09-252-991A-13657
 : Sequence 13657, Application US/09252991A
 : Patent No. 6551795
 : GENERAL INFORMATION:
 : APPLICANT: Marc J. Rubenfield et al.
 : TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 : TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 : FILE REFERENCE: 107196, 136
 : CURRENT APPLICATION NUMBER: US/09/252,991A
 : CURRENT FILING DATE: 1999-02-18
 : PRIOR APPLICATION NUMBER: US 60/074,788
 : PRIOR FILING DATE: 1998-02-18
 : PRIOR APPLICATION NUMBER: US 60/094,190
 : PRIOR FILING DATE: 1998-07-27
 : NUMBER OF SEQ ID NOS: 33142
 : SEQ ID NO 13657
 : LENGTH: 2175
 : TYPE: DNA
 : ORGANISM: Pseudomonas aeruginosa
 : US-09-252-991A-13657

	Query Match	2.3%	Score 54	DB 4	Length 2175	
	Best Local Similarity	45.3%	Pred. No. 8.3e-05			
	Matches 250	Conservative 0	Mismatches 290	Indels 12	Gaps 1	
QY	299	ACCAAGCGAGATCCTCAAGATGTCTCCTTGTACGTGTGAGAGCGGGACATCATGTGCATC	358			
Db	1489	AGCCCGCGGGTGTGTGACGTGCACAGGCCCTTGAAACATCCCGCGCGGACGGGTGGCGCTG	1541			
QY	359	CTAGGAAGCTTCAGGCTCCGGGAAAACCGGTGTGCGAGCGGCATGCTCCGGGAGCGCTGGGG	418			
Db	1549	CTGGGGCGGCMAAGGCGCGCGGCACATGTCAGACCTTGCTGTCTGTAGCGGCGCTGTGTCAGC	1601			
QY	419	CGCGCGGGGACCTTCTCTGGGGGAGGTGTATGTGAACGGCGGGGCGCTGCGCGGGAGCAG	478			
Db	1609	GCGCAGCGGGACGCGCTGTGCTCTGTGAGACACTCAGCCATGACCCAGATGACCCGGGCGGAC	1661			
QY	479	TTCCAGAGACTGCTTCTCTCTACGTCTGTGAGAGCACAACCTGTGAGAGCGCTTCAACCGTG	538			

Db 1669 CGCAGAGCGGTATCGGCTACCTGCCGAGAGCGTGGCGCTGTTCCATGGACG----- 1722

OY 539 CGCGAGAGCTGCATACACCGCGCTGCTGGCCATCCGGCGGCAATCCCGCTCTTC 598

Db 1723 -----CTGCGGACAACTCAACCTGAGAAAGCGCCCTGGGCGACAGAGAACTGCTG 1776

OY 599 CAGAGAGGTGAGGCGCTCATGSCAGAGCTGAGTCTGAGCCATGTGGCAGACCGACTG 658

Db 1777 GAGACCTGAGCGGGGTCCGGCTGGCGGCTTCGTCGGCGGCAACCGCTGGGGCTGGAC 1836

OY 659 ATTGGCACTACAGCTGGGGGGGATTCCAGCGGGGAGCGCGCGGCTTCATCGCA 718

Db 1837 ATGCCATTCAGAGGCGACCGCAGCTGCGGCGGCAACCGCGCTCGGGCTGGCC 1896

OY 719 GCCCAGCTGCTCAGAGATCTTAAGTCATGCTTTGATGAGCCAAACGACGCTGGAC 778

Db 1897 CGGTCCTCTACAGGACCTCGATCCCTGCTGCTGAGAGCGACCGCGGCTTCGAC 1956

OY 779 TGCATGACTGCTAATCAATGTTGCTCTCTGTTGGAAGTGGCTCGAGAAACGAATT 838

Db 1957 CAGGCGAGGAGAAAGATGATGACTAGCTGAGCAATGTTGGCAAGCGCACCTTG 2016

OY 839 GTGTTCTCACC 850

Db 2017 GTCATCACCAAC 2028

RESULT 13

US-09-252-991A-13575

; Sequence 13575, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 13575

; LENGTH: 2292

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-13575

Query Match 2.3%; Score 54; DB 4; Length 2292;

Best Local Similarity 45.3%; Pred. No. 8.6e-05;

Matches 250; Conservative 0; Mismatches 290; Indels 12; Gaps 1;

OY 299 ACCAGGAGATCTCAAGATGTTCTCTGTACGTGAGAGCGGAGATCATGTGCATC 358

Db 1460 AGCCCGCGGTGTACACGTGACGGCTCTGACATCCGGCGGAGCGGGTGGGCTG 1519

OY 359 CTAGGAAGCTCAGGCTCCGGGAAACACGCTCTGAGCGCATGTCGGGAGGCTGGG 418

Db 1520 CTGGGCGGCAACGCGCGGCAAGTGCACCTCTGCTGCTGACGCGGCTGCTGAC 1579

OY 419 CGCGGCGGACCTTCTGGGGGAGTGTGTGAAGGGCGGGGCTGGCGGGAGAG 478

Db 1580 GCGCAGCGGAGAGCGCTGCTGCTGAGAGCTGACCTTACACCAAGTTCAGCGCGGAC 1639

OY 479 TTCAGAGACTGCTTCTCTAGTCTGAGAGCGACACCTCTGAGACGCTTCACCGT 538

Db 1640 CGCCACCGGATGATGCTGCTGCGGAGAGCGGCGCTTTCATGAGCAGC----- 1693

OY 539 CGCGAGAGCTGACACCGCGCTGCTGCGCATCCGCGCGCAATCCGGCTCTTC 598

Db 1694 -----CTGCGGCAACCTCAACTGGAGAAAGCGCGGCTGGGCGAGAGAACTGCTG 1747

OY 599 CAGAGAGGTGAGGCGCTCATGAGCAGAGCTGAGTCTGAGCCATGTGGCAGACGACTG 658

Db 1748 GAGACCTCGACGAGGGGTGCGCTGGCGGCTTCTGCGGGCCACCCCGCTGGGCTGAC 1807

OY 659 ATTGGCACTACAGCTGGGGGGGCAATTCACGGGTGAGCGCGCGGCTTCATCGCA 718

Db 1808 ATGCCATTCAGAGGCGACCGCAGCTGTCGGCGGCAACCGCAGGCGCTCGGGCTGGCC 1867

OY 719 GCCCAGCTGCTCAGAGATCTTAAGTCATGCTTTGATGAGCCAAACGACGCTGGAC 778

Db 1868 CGGTCCTCTACAGGACCTCGATCCCTGCTGCTGAGAGCGACCGCGGCTTCGAC 1927

OY 779 TGCATGACTGCTAATCAATGTTGCTCTCTGTTGGAAGTGGCTCGAGAAACGAATT 838

Db 1928 CAGGCGAGGAGAAACAGATGATGACTAGCTGAGCAATGTTGGCAAGCGCACCTTG 1987

OY 839 GTGTTCTCACC 850

Db 1988 GTCATCACCAAC 1999

RESULT 14

US-09-252-991A-10998

; Sequence 10998, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 10998

; LENGTH: 993

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-10998

Query Match 2.3%; Score 53.4; DB 4; Length 993;

Best Local Similarity 48.3%; Pred. No. 8e-05;

Matches 249; Conservative 0; Mismatches 251; Indels 15; Gaps 3;

OY 306 AGATCTCAAGATGTTCTCTGTACGTGAGAGCGGCGAGATCATGTGCATCTAGGA 365

Db 50 AGCCCTCGAGCAACATCACTGATATCCAGAGCGGCGAAGCTGTGCGCTCTGCGC 109

OY 366 GCTCAGGCTCGGGAACACGCTCTGAGACGCTGTCGGGAGCTGGGCGCGG 425

Db 110 CTTCCGGGTGCGGAAGACACCTCTGCTGATCAATCCGCG---GGCTGAAACCCGG 166

OY 426 GGAATCTTCCGCGGAGGTGTATGTAAGCGCGCGGCGCTGCGCGCGGAGCACTTCAG 485

Db 167 AGCCCGGCAACATCTGTTTCATATGCGAGAGAGCTCTGCGAGCAGAGCTC---GCGACC 223

OY 486 ACTGCTTCTCTAGCTCTGAGAGGAGACACCTCTGAGACAGCCCTCACCGTGGGAG 545

Db 224 GCAACGTGCGGCTTCTTCAAGACTATGCGCTGTTCCGCACATGAGAGGTGTTGACA 283

OY 546 CGTGTACTACACCGGCTGCTGCGCATCCGCGGCGCAATCCCGCTCTTCAGAGAGA 605

Db 284 AGTGCCTTTCGCGCTGCGCATGAAGCCCAAGGGGAGCGGCGGCGAGTCCGGGATCA 343

OY 606 AGTGTAGGCGCTCATGCGAGAGCTGAGTCTAGCGCATGTGACAGACGACTGATTTGCA 665

Db 344 AGCGCAAGTTCACAGAGCTG-----CTCAACATGTGTGAGCTGCACTGCTGCGG 394

OY 666 ACTACAGCTTGGGGGCAATTCACGAGGTAGCGGCGCGGCTGCTCATCGAGCCAGC 725

Db 395 ACCGCTATCCAGACCACTCTCGGCGCAGCGGACGATATCGCCCTGGCCGCGCGC 454
OY 726 TGCATCAGGATCTTAAGTCTATGCTGTGATGAGCCACACAGCCCTGGACATGATGA 785
Db 455 TGGCGGTGAGCCGAAATCTCTCTCTGACGAACTTCGCGCCCTCGACGCGCAAG 514
OY 786 CTGCTAATCAGATTGCTCTCTCTGATGAGTGA 820
Db 515 TACGCAAGAGACTCGCGCGCTGCGCTGCGCGCGCTG 549

RESULT 15

US-09-252-991A-10934
: Sequence 10934, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 3142
: SEQ ID NO 10934
: LENGTH: 1335
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10934

Query Match 2.3%; Score 53.4; DB 4; Length 1335;
Best Local Similarity 48.3%; Pred. No. 9,4e-05;
Matches 249; Conservative 0; Mismatches 251; Indels 15; Gaps 3;

OY 306 AGATCTCAAGATGCTCTTGTACGTGAGAGCGGCGACATCATGTGATCTTAGAA 365
Db 27 AGGCCCTGGACACATCATGATATCCAGAGCGCGAATGTGTGCGCTCGCGCC 86
OY 366 GCTCAGGCTCCGGGAAACAGCGTGTGAGACCGCATGTCCGGAGCTGGGGCGCGC 425
Db 87 CTTCCGGGTGGGAAAGACACCTGCTGCGATCATGCGCG--GGCTGGAACCCCGG 143
OY 426 GGAACCTTCCTGGGGAGTGTATGTAACCGCGCGCTGCGCGGAGCAGTTCAGG 485
Db 144 ACGCGGCAACATGTTTCATGCGGAGACGCTTCGACAGCAGAGCTGC--GCGACC 200
OY 486 ACTGCTTCTCTACGTCTTCAGAGGACACCTGCTGAGCAGCTTCACCGTGGCGAGA 545
Db 201 GCAACGTGCGCTTCCTTCACACATGACGCTGTCCGCGACATGACGCTGTTCGACA 260
OY 546 CGCTGACATACCGGCGCTGCGCATCCGCGCGCATCCCGGCTCTCTCCAGAGA 605
Db 261 ACGTCCCTTCGCGCTGCGCATAGCCCAAGGCGGCGCGGCGAGTCCGCGATCA 320
OY 606 AGGTGAGCGCGTATGAGCAGAGCTGAGCCATGTGCGAGACCGAGCTGATTTGCA 665
Db 321 AGGCCAAGTTCAGAGCTG-----CTCAACATGTGCGAGCTGACTGGCTCGCCG 371
OY 666 ACTACAGCTGGGGGCGATTTCCACGCGGTGAGCGCGCGGCTCTCCATCGACGCCAGC 725
Db 372 ACCGCTATCCAGACACTCTCGGCGCGGCGAGCGATATGCGCTGCGCGCGCG 431
OY 726 TGCATCAGGATCTTAAGTCTATGCTGTGATGAGCCACACAGGCTGGACTGCATGA 785
Db 432 TGGCGGTGAGCCGAAATCTCTCTCTGACGAACTTCGCGCGCTTCGACGCGCAAG 491
OY 786 CTGCTAATCAGATTGCTCTCTCTGATGAGTGA 820
Db 492 TACGCAAGAGCTCGCGCGCTGCGCGCGCTG 526

Search completed: July 27, 2003, 12:45:40
Job time: 165.622 secs